Engineered DNA-Binding Proteins for Targeted Genome Editing and Gene Regulation

Engineered DNA-Binding Proteins for Targeted Genome Editing and Gene Regulation

A dissertation presented

by

Morgan Lee Maeder

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of Genetics

Harvard University

Cambridge, Massachusetts

May 2013

Dissertation Advisor: Dr. J. Keith Joung Morgan Lee Maeder

Engineered DNA-Binding Proteins for Targeted Genome Editing and Gene Regulation

Abstract

Engineered DNA-binding proteins enable targeted manipulation of the genome. Zinc fingers are the most well characterized DNA-binding domain and for many years research has focused on understanding and manipulating the sequence-specificities of these proteins. Recently, major advances in the ability to engineer zinc finger proteins, as well as the discovery of a new class of DNA-binding domains - transcription activator-like effectors (TALEs), have made it possible to rapidly and reliably engineer proteins targeted to any sequence of interest. With this capability, focus has shifted to exploring the applications of this powerful technology. In this dissertation I explore three important applications of engineered DNA-binding proteins.

The ability to specifically modify the genome has great potential for both research and therapeutic applications. While zinc finger nuclease (ZFN)-induced double-strand breaks had been shown to stimulate homology-directed repair (HDR) in a variety of cell types and organisms, this technique had never been used to correct a disease-causing mutation in human induced pluripotent stem cells (hiPSCs). I used ZFNs engineered to the human β-globin locus to correct the sickle cell anemia mutation in induced pluripotent stem cells (iPS cells) derived from patients with this disease.

Another important application of DNA-binding proteins is the construction of site-specific transcription factors. I performed a systematic, large-scale analysis of protein variants

that defined an optimal framework for making monomeric TALE-based transcriptional activator proteins. I then used this platform to demonstrate that TALE-activators can function robustly and synergistically to activate transcription of target genes.

Having determined that monomeric TALE fusion proteins can function robustly as activators, I next explored whether TALEs could be used to target the enzymatic activity of the DNA hydroxymethylase TET1, a protein important for demethylation of methylated CpGs. I engineered novel TALE-TET1 fusions and demonstrated that these proteins could be used to direct demethylation activity to specific target CpGs in living cells.

Each of these studies examines a different potential application of customized DNA-binding domain technology and lays important foundations for future work in these fields.

Table of Contents

Acknowledgements

First and foremost, I owe my deepest gratitude to Keith for his enthusiasm, patience, expertise and support. Thank you for always having confidence in me and for putting up with me. For introducing me to Pinocchio's Pizza and for saving me from going to med school. Thank you for your guidance and mentorship, for pushing me to work harder than I thought I could, and for always believing that I could succeed. I could not have asked for a better advisor.

Thank you to all the Joung lab members, past and present. Spending all my time in the lab would not have been possible if you didn't make it such a fun place to be. Thank you to Mat for your organization and unending patience, and to James, Sam and Vinny for being the most amazing techs. Thank you to Shengdar, Deepak and Jeff for being not only my lab mates, but also my best friends. Specifically, thank you to Shengdar for all your help analyzing data and for always pointing out the obvious. To Deepak for reminding me not to take things too seriously, your wonderful DJ skills and for always knowing when it's time for a beer. And to Jeff for teaching me to drink whiskey, always being there when I need someone to talk to, constant advice and help with experiments, Saturday morning bloody marys, and for having the patience to sit next to me for the last four years.

Thank you to all my science teachers over the years. To Dr. Dorfman for introducing me to Biology, Dr. Ferguson for giving me my first lab experience and Dr. Eastman for making me fall in love with research.

And of course, thank you to my family for all their love and support. To my parents for instilling in me a love of learning and a desire for academic excellence. To my siblings for always being there. And to David, for everything. For putting up with long hours spent in the lab, for being interested in my work even when you don't understand it, and for loving me, not in spite of, but because of my dorkiness.

To Keith

For believing in the technology and for believing in me.

Chapter 1

Introduction:

Engineered DNA-Binding Proteins and Their Applications

For Targeted Genome Editing and Gene Regulation

Morgan L. Maeder

Technologies for performing targeted manipulation of genetic sequence or gene expression are powerful methods with broad applications in research and therapeutics. The most direct strategy for studying the regulation and function of a gene is to directly manipulate its DNA sequence in its endogenous context. Gene knockouts provide one powerful approach for elucidating gene function, but the ability to more precisely alter genetic sequence would also enable more subtle studies, such as SNP modeling or tagging an endogenous gene product to follow its expression and localization. Precise modification of the genomes of a wide variety of species would enable powerful reverse genetics approaches to discerning gene function and make it possible to recapitulate human disease phenotypes in other organisms. In addition, the ability to alter gene expression in a targeted fashion would also allow for in-depth studies into gene function and regulation in different developmental and disease states. In the longer-term, targeted manipulation of the genome and of endogenous gene expression both hold great promise for therapeutics. The ability to correct disease-causing mutations or alleviate symptoms by specifically repressing a disease allele or up-regulating a compensatory gene could enable therapeutic strategies that would directly address the root cause of gene-based diseases.

Engineered DNA-binding proteins can be used to confer sequence specificity to a wide variety of effector proteins. When fused to nuclease domains, such chimeric enzymes can induce site-specific double strand DNA breaks, which can mobilize DNA repair mechanisms in living cells that can be exploited to introduce changes in DNA sequence with high efficiency. This method has been used to introduce targeted mutations and insertions in model organisms and human cells and has the potential to serve as a gene therapy approach for genetic

disorders. Additionally, fusions to transcriptional activators or repressors can be used to specifically modulate the expression of target genes. The ability to alter gene expression in a targeted fashion enables in-depth studies of gene function as well as manipulation of genetic pathways to achieve desired phenotypes. These approaches rely on the ability to engineer site-specific DNA binding domains and here I review the several platforms that have been developed over the past two decades for this purpose.

Engineered DNA-binding Domains

Zinc Fingers

Zinc finger (ZF) proteins are the most abundant class of human transcription factors and the Cys_2-His_2 zinc finger DNA binding domain is found in approximately 2% of human genes[1]. The Cys_2-His_2 zinc finger domain was originally identified as a repetitive sequence in the TFIIIA gene of *Xenopus laevis* with the consensus sequence X_2-C-$X_{2,4}$-C-X_{12}-H-$X_{3\text{-}5}$-H where X represents any amino acid[2]. In the presence of a zinc atom, this domain forms a compact $\beta\beta\alpha$ structure[3,4]. The crystal structure of Zif268 bound to DNA has served as the basis for our understanding of DNA binding by zinc fingers[4,5]. The α-helical portion of each finger fits

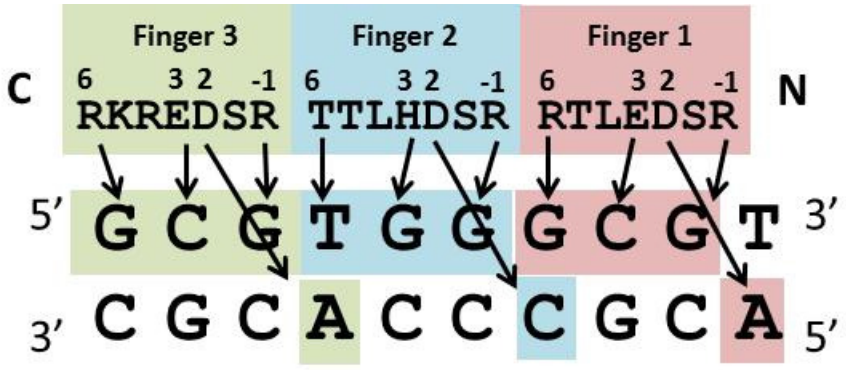


Figure 1.1. DNA Recognition by Zif268. Schematic indicates base contacts (depicted with arrows) made from positions -1, 2, 3 and 6 of the α-helices. Protein is shown C-terminal to N-terminal and DNA 5' to 3'. Figure adapted from Wolfe, Nekludova and Pabo, *Annu. Rev. Biophys. Biomol. Struct*. 2000.

into the major groove of the DNA. Major DNA contacts are made by amino acids at positions -1, 3 and 6 of the α-helix and the amino acid at position 2 may sometimes make a "cross-strand" contact to the complementary DNA strand. Tandem fingers in the Zif268 protein wrap around the DNA to bind overlapping four base pair sub-sites such that this three-finger protein recognizes 10 bps of DNA (Figure 1.1)

The apparently modular structure of Zif268 suggested that these proteins might provide an attractive framework for engineering novel DNA-binding motifs. Initial attempts to design ZFs with novel specificities based on a simple set of rules met with some success[6,7], however, these design experiments, as well as additional structural studies, clearly indicated that different ZF proteins can have different docking arrangements on the DNA and that there are significant context-dependent effects occurring between fingers that could not be easily predicted by a simple recognition code[8-10]. The limited success of designing ZFs based on a simple recognition code led researchers to explore the use of selections from combinatorial libraries of ZFs, an approach that would enable identification of proteins with novel specificities without the need to comprehensively understand all aspects of zinc finger-DNA recognition.

Combinatorial libraries combined with selection-based methods proved to be a more robust approach than code-based design approaches for generating individual fingers with novel DNA-binding specificities. The phage display method involves displaying zinc finger proteins on the surface of filamentous phage and performing sequential rounds of affinity selection with biotinylated target DNA to enrich for phage expressing proteins able to bind the specific target sequence[11]. By randomizing key residues of one finger of Zif268 and holding the other two fingers constant to serve as anchors over the binding site, initial experiments

demonstrated that phage display could interrogate such large libraries and select for variants with novel DNA-binding specificities[11,12 13,14]. The bacterial-two-hybrid (B2H) system has been adapted to allow for similar selection of zinc finger proteins that bind specific target sites from randomized libraries [15]. In this system, the ZF binding site is placed upstream of a weak promoter driving expression of two selectable markers in *Escherichia coli*. A library of ZFs, fused to a fragment of the yeast Gal11P protein, are expressed in the cells and binding of a zinc finger to the target site recruits an RNA polymerase-Gal4 fusion, thus activating transcription and allowing survival of the cells on selective medium. Both phage display and the B2H system enabled the engineering of single fingers with novel DNA-binding specificity; however, the libraries necessary to select an entire three-finger array by these systems are prohibitively large since all three fingers would need to be simultaneously randomized to best account for context-dependent effects between and among fingers in the array. The next challenge in the field was therefore to develop technologies that would enable the engineering of multi-finger arrays.

The simplest method of engineering multi-finger ZFs, known as "modular assembly" relies on archives of single fingers with pre-selected specificities and the theory that these "modules" can be joined together to form longer, functional arrays (Figure 1.2A). Klug and colleagues initially used modular assembly to engineer a novel three-finger protein targeted to the BCR-ABL fusion oncogene[16] and several groups have developed collections of single-finger modules, either identified in naturally occurring proteins[17] or selected to bind specific three base pair target sites[18-21], which can be used to practice modular assembly. While this method is easy to practice, it does not take into account the well-documented context-dependent

interactions of fingers within a multi-finger array[9,22-24] and several studies have shown that there is a high failure-rate for proteins engineered by modular assembly[25,26].

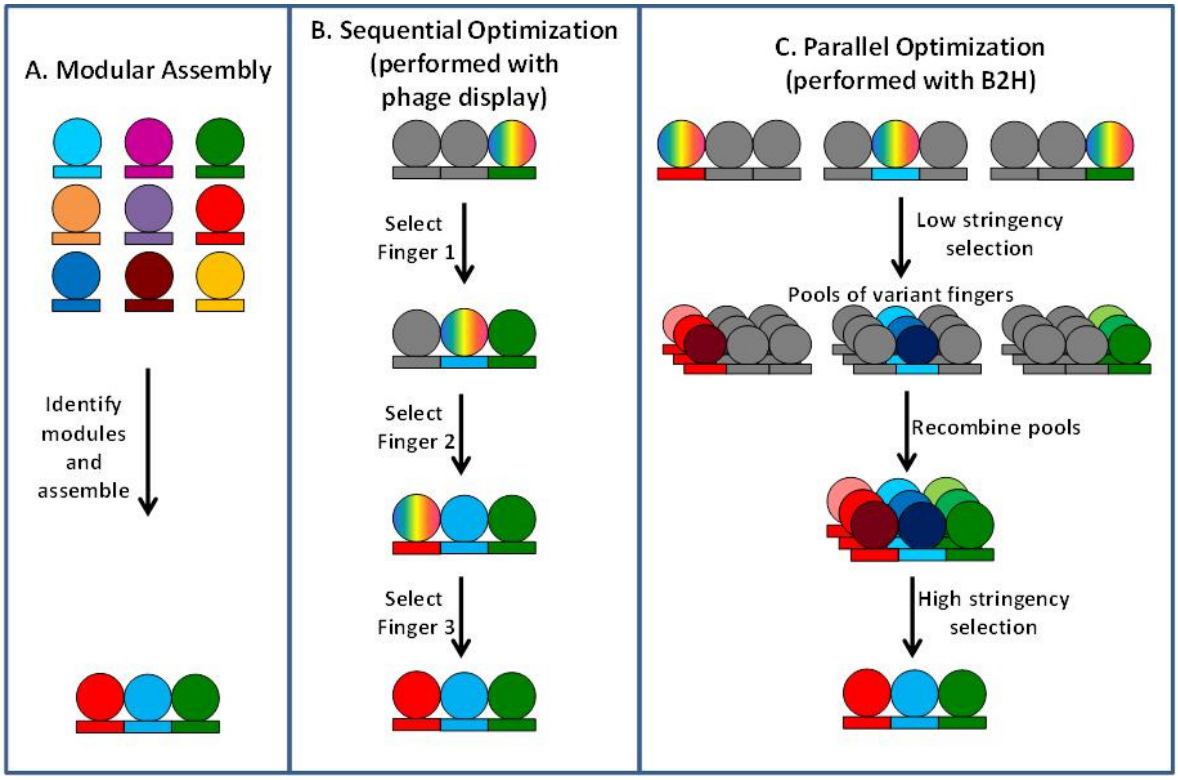


Figure 1.2. Methods of engineering multi-finger ZF arrays. Schematics display three methods for engineering ZF arrays: (A) modular assembly, (B) sequential optimization performed by phage display (as described by Greisman and Pabo, *Science* 1997) and (C) parallel optimization (performed using the B2H as described by Hurt et al., *PNAS* 2003 and Maeder et al., *Mol Cell* 2008). Figure adapted from Thibodeau and Joung*, Discov Med* 2003.

Using a sequential selection strategy in which fingers were serially added and optimized, Greisman and Pabo were able to obtain novel three-finger proteins targeted to a TATA box, a p53 binding site and a nuclear receptor element (Figure 1.2B)[27]. In this method, the first finger of a three-finger protein is randomized and selected by phage display and a small pool of selected fingers is carried forward to the next stage in which a second finger is randomized and selected. This process is then repeated for the third finger. By utilizing a pool of variants at each step, it is possible to select for combinations of fingers that function together to bind the

target site, and thus account for the context dependent interactions between fingers. While this scheme enables selection of multi-finger arrays that bind their target sites with high affinity and specificity[22], there are several limitations that inhibited broad adoption of this method. Each stage of the selection requires construction of a new, large combinatorial library incorporating the pool of selected fingers from the previous round. Additionally, multiple rounds of selection and amplification are required for each stage of the phage display protocol and this must then be repeated for each successive finger of the final protein. With five to eight rounds of phage display enrichment required for selection of each of the three-fingers, approximately 15-24 iterative rounds of phage display are necessary to engineer a three-finger protein.

Joung and colleagues developed a strategy of parallel optimization to allow for selection of multi-finger arrays without the need for rounds of consecutive selections (Figure 1.2C)[28]. In three initial parallel selections, master randomized libraries were interrogated using the B2H system under low selection stringency to identify a variety of individual fingers capable of binding each 3 base pair sub-site of the target site. These three selected populations were then randomly shuffled to generate a library of multi-finger proteins, which was interrogated under high-stringency selection conditions to identify three-finger proteins targeted to a specific nine base pair site. This method successfully generated highly specific proteins targeted to a *BCR-ABL* translocation sequence, a site in the *ERB2* gene and a site in the HIV promoter and characterization of these proteins demonstrated that they bound their target sites with higher affinities and specificities than proteins engineered by modular assembly for the same target sites[28,29]. The parallel optimization method, like sequential selection, accounts for context-

dependent effects between neighboring fingers because pools of individual fingers are carried forward from the first round of selection, allowing for final selection of combinations of fingers that work well together in the context of a three-finger protein. Unlike sequential selection, however, this method relies on three master randomized libraries for initial selection against 3bp target sites, and only requires construction of a single targeted library for each full target site. Removing the requirement for repeated construction of libraries and iterative rounds of selection makes this method dramatically faster and easier to practice than the sequential selection approach.

To further simplify and enable the broad adaption of this method, Joung and colleagues went on to perform initial low-stringency selections for a large number of three base pair subsites and made the resulting targeted libraries, each consisting of approximately 95 different proteins, available to the public[30]. Termed oligomerized pool engineering (OPEN), this method now allowed for selection of novel three-finger proteins simply by recombining the pre-existing single-finger libraries to generate a targeted library of approximately one million arrays and interrogating this library using the B2H system to identify proteins that bind with high affinity and specificity to any target site of interest. Proteins selected using this method have been shown to function in human somatic and pluripotent stem cells[30-33], plants[30,34,35] and zebrafish[36].

Analysis of thousands of proteins selected by the OPEN method suggested a selection-free method for assembling novel three-finger proteins that at least partly accounts for the context-dependent interactions among fingers. By identifying two different three-finger arrays with a common middle finger, it is possible to recombine the N- and C-terminal fingers from the

two different arrays to generate novel proteins while maintaining interactions of these end fingers with the neighboring middle finger[37]. This selection-free strategy, known as context-dependent assembly (CoDA), enables rapid construction of ZF arrays with novel binding specificities that function efficiently in zebrafish and plants. However, due to its reliance on archives of three-finger proteins with certain defined characteristics (i.e.—shared, common middle fingers), the targeting range of CoDA is more limited than OPEN.

The proprietary engineering method of Sangamo Biosciences relies on assembling ZFs from a pre-existing archive of two-finger proteins, each of which bind a six base pair target site[38,39]. The technology for this method derives from the finding that two-finger units could be joined together by non-canonical, flexible linkers to generate ZF arrays with high target specificity[40]. While this method has been used to generate many highly functional proteins[39,41-43], it relies on access to a validated proprietary library of pre-selected proteins and is therefore not a viable, publically available method. This system has been commercialized by Sigma-Aldrich but for many years the price was high enough to be out of the range of many researchers[44].

Transcription Activator-Like Effectors Repeat Domains

The discovery that highly conserved repeat domains from transcription activator-like effectors (TALEs) of the plant pathogen species *Xanthomonas* are DNA-binding domains whose specificity is governed by a simple one-to-one code again raised the exciting possibility that these proteins might be amenable to the modular assembly method of engineering that had proven unsuccessful for zinc fingers[45,46]. Highly conserved 33-35 amino acid TALE repeats each

bind a single base pair of DNA with specificity dictated by two hypervariable residues (Figure 1.3). The simple relationship between the identity of these hypervariable residues and the DNA base bound by the repeat was initially revealed by Boch and colleagues and they validated this code by engineering the first synthetic TALE arrays with novel DNA-binding specificities[45]. Further confirmation was provided by computational analysis of naturally-occurring TALEs[46] and by additional engineering of TALE proteins with novel specificities[47-50]. Crystal structures of TALEs bound to DNA revealed that each repeat forms a two-helix structure connected by a loop containing the hypervariable residues, which are presented into the major groove of DNA as the protein wraps around the DNA in a superhelical structure[51,52]. Of the two hypervariable residues, only the one at position 13 makes direct base-specific contacts with the DNA, while residue 12 serves to stabilize the conformation of the loop through hydrogen bonding, thus ensuring correct presentation of the loop into the major groove of the DNA and allowing residue 13 to specifically recognize DNA bases.

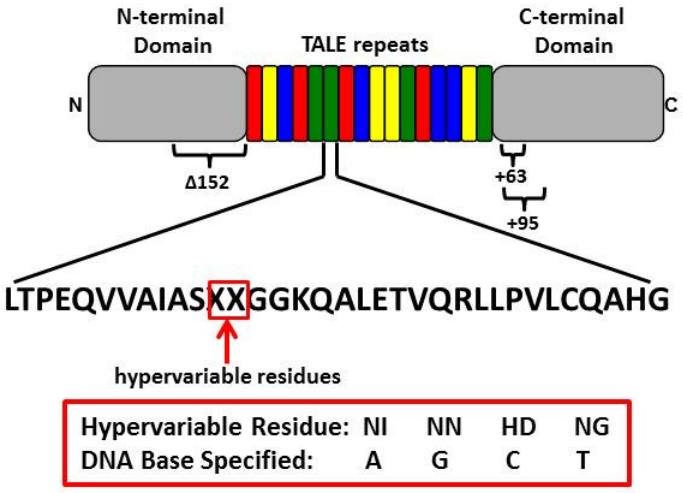


Figure 1.3. Structure of a TALE. Schematic illustrates the structure of a naturally occurring TALE. The commonly used Δ152 N-terminal truncation and +63 and +95 C-terminal truncations are shown. The sequence of a single TALE repeat is shown with hypervariable residues boxed in red and the four most commonly used amino acid pairs and corresponding DNA bases specified shown below.

The efficient binding of TALE repeat arrays also requires additional flanking N- and C-terminal amino acid sequences (Figure 1.3). Initial attempts to engineer custom TALE proteins

involved systematically truncating these domains to identify the minimal sequence necessary to maintain DNA-binding function. These deletion experiments revealed that removal of 141, 152 or 153 amino acids from the N-terminal domain did not dramatically reduce TALE activity and that over 100 amino acids of the N-terminus were necessary for DNA binding[48,53,54]. The crystal structure of a TALE containing this minimal N-terminal region showed that it contains four atypical repeat domains with high structural similarity to the canonical TALE repeats and which are necessary for DNA binding[55]. C-terminal to the TALE repeats is a half repeat, referred to as the "0.5 domain", which also binds a single base pair of DNA, a nuclear localization signal and an endogenous activation domain. While initial studies incorporated large portions of the endogenous C-terminus domain[49], further work demonstrated that truncations of this domain could lead to increased activity of engineered TALEs[48,53,56]. Unlike the N-terminus, less of a consensus has been reached regarding C-terminal truncations and several different deletions can be used to generate functional TALEs. Additionally, it appears that the optimal length of the C-terminal domain may be different depending on the functional domain fused to the engineered TALE[48].

Several platforms exist to enable engineering of TALE arrays and these can be broadly grouped by the throughput of the method[57]. The simplest and lowest throughput methods use standard cloning techniques to assemble TALEs from archives of plasmids, each consisting of single TALE repeats[58,59]. A medium throughput version of one of these methods also relies on standard restriction digestion and ligation, but uses an archive of pre-assembled multi-repeat TALEs as a starting point, thus enabling assembly of a full length TALE array in far fewer steps, with the downside of requiring a larger number of plasmids (over ten times more) than the

single-repeat version[60]. Several other medium-throughput assembly methods rely on the "Golden Gate" system, which combines digestion of multiple plasmids with type IIs restriction enzymes and ligation of the resulting fragments in a single reaction[50,61-66]. These methods are relatively simple to implement, however, many of them have limitations in the length of TALE arrays that can be constructed. Another class of medium-throughput methods relies on solid-phase assembly. The iterative capped assembly (ICA) method uses sequential ligation of single-repeat fragments and capping with a hairpin oligo to prevent extension of incomplete chains[67] while another strategy uses ligation of pre-assembled four-repeat units[68]. The Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) assembly method also involves iterative ligations of four-mers on a solid-phase support and is amenable to either medium-throughput if practiced manually or high throughput assembly if automated on a liquid-handling robot[69]. Using automated FLASH, up to 96 TALE arrays can be assembled in a single day. A recently described high-throughput Golden-Gate method enables assembly of TALEs consisting of 14.5 to 18.5 repeats in a single reaction, and was used to generate 126 functional TALEN pairs[70]. An alternative high-throughput method uses ligation independent cloning (LIC) for high precision assembly five-mer TALE repeats, however, this method requires an archive of 3,072 plasmids and can only be used to assemble 15.5-mer TALEs[71]. In addition to the speed of assembly, these various platforms differ in the TALE architecture used, the efficiency of the method and the length of arrays that can be constructed. Customized engineered TALE arrays are now also available through several commercial vendors.

Targetable Gene-Editing Nucleases

Genome Editing

Since the development of recombinant DNA technology in the 1970s, the concept of gene therapy – modifying the genome to treat diseases caused by genetic mutation – has been a focus in the field of biomedical research [72]. Foundational to the current state of the field was the discovery that site-specific double-strand breaks (DSBs) could be used to stimulate endogenous cellular repair machinery [73-76]. Double strand breaks in DNA are typically repaired by either non-homologous end joining (NHEJ) or homology-directed repair (HDR)[77]. NHEJ, the predominant repair pathway, functions by re-ligating the cleaved ends of DNA in the absence of a homologous donor sequence[78]. This pathway is error-prone and can often result in insertions and/or deletions (indels) at the site of the break (Figure 1.4A). Taking advantage of these indels, stimulation of NHEJ by targeted DSBs has been used to knockout target genes in a wide variety of cell types and organisms, such as *Drosophila*, plants, zebrafish and rats[39,79-82].

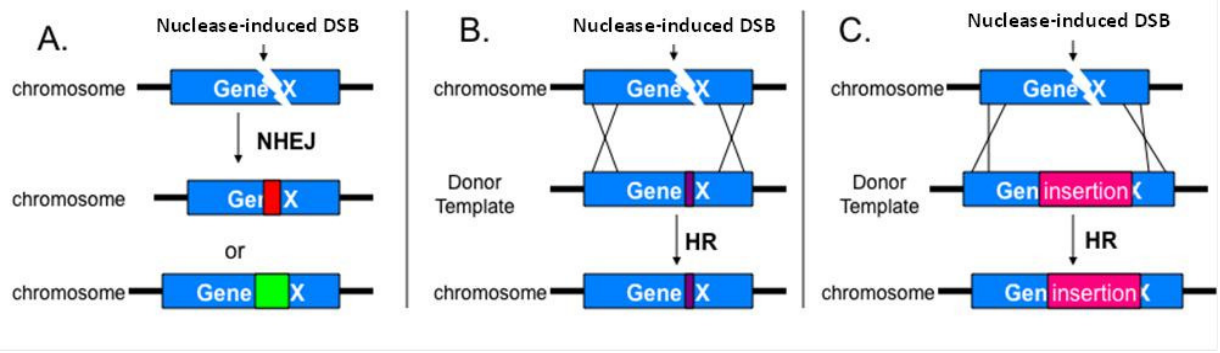


Figure 1.4. Nuclease-Mediated Genome Modifications. Targeted nucleases introduce site-specific double strand breaks that induce **(A)** non-homologous end joining, which leads to mutagenesis by introduction of insertions (green) or deletions (red), **(B)** homologous recombination with a donor template bearing a point mutation (purple) or **(C)** homologous recombination with a donor template containing an insertion (pink).

HDR relies on strand invasion by a homologous sequence of DNA and subsequent repair of the break in a template-dependent manner[83]. Initial experiments with the meganuclease I-*SceI*

demonstrated that introduction of a DSB increased the rate of homologous recombination with an exogenous donor DNA by several orders of magnitude[73-75]. Donor constructs were constructed by flanking a desired insertion or mutation with sequence homologous to the target site. Upon DSB induction, the cell may use this donor construct as a repair template and thus incorporate the desired change into the endogenous locus. This method can be used to make specific point mutations or to insert larger cassettes, such as transgenes or tags (Figure 1.4B and C). While plasmids are traditionally the most commonly used source of donor DNA, recent studies have shown that single stranded oligonucleotides (ssODNs), with as little as 80 base pairs of homology, can serve as efficient donor templates for HDR[56,84,85]. Nuclease-induced HDR has been shown to function in a wide variety of cells and organisms such as *Drosophila*, plants, zebrafish and human somatic and pluripotent stem cells[30,31,37,41,56,58,86-88].

Because DSB-induced gene editing relies on the endogenous repair mechanisms of the cell, it is universally applicable to any cell type or organism that employs these methods for DNA repair. Several platforms for inducing site-specific DSBs have been described over the past 17 years: zinc finger nucleases (ZFNs) and TALE-nucleases (TALENs) are constructed by fusing a non-specific nuclease domain to engineered DNA-binding domains, meganucleases involve re-engineering naturally occurring homing endonucleases to recognize novel target sequences and the recently described CRISPR/Cas system relies on RNA-guided recruitment of the Cas9 nuclease to DNA target sites (Figure 1.5).

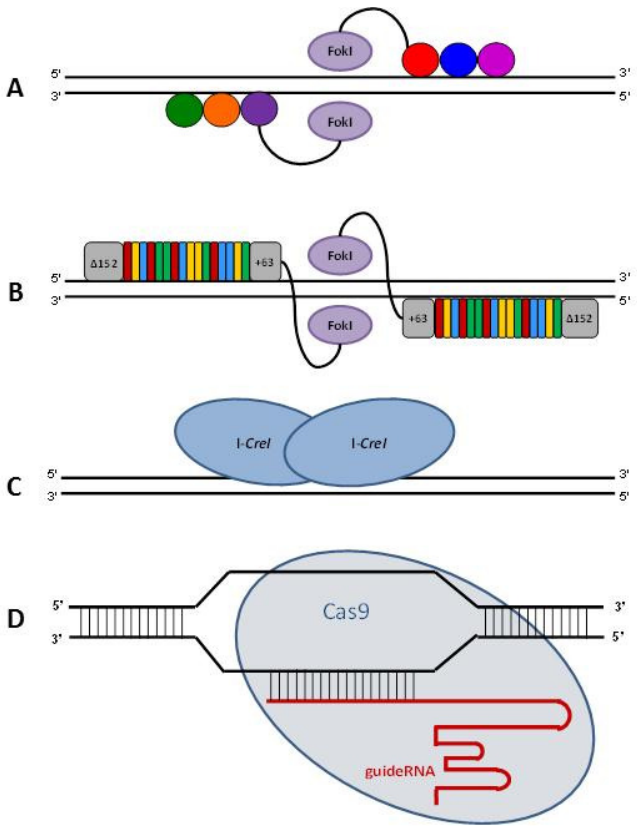


Figure 1.5. Targeted Nuclease Platforms. Schematics illustrate the basic architecture of A) zinc finger nucleases, B) TALE nucleases, C) meganucleases and D) CRISPR RNA-guided Cas9 nuclease.

Zinc Finger Nucleases

Zinc Finger Nuclease (ZFN) technology was made possible by the realization that the DNA-binding domain and cleavage domain of the FokI restriction endonuclease function independently of one another[89]. Therefore, it is possible to replace the DNA-binding domain with a zinc finger domain to generate a chimeric nuclease with novel binding specificity[90,91]. The FokI nuclease functions as a dimer and induction of a DSB requires two ZFNs, each consisting of a zinc finger domain fused to a FokI monomer, that bind opposite strands of DNA and are separated by 5-7bp.

Initial experiments performed in *Xenopus* and *Drosophila* showed that ZFN-induced DSBs could be used to modify the genome through either NHEJ or HDR[79,86] and soon after it was

shown that ZFNs could stimulate HDR of a GFP reporter in human cells[92]. Subsequently, this technology has been applied to endogenous genes in human somatic[41 30,87]and pluripotent stem cells[31,43,88,93,94], as well as to a wide variety of whole organisms, such as zebrafish[36,37,39,82], plants[34,95,96], rats[81], mice[97] and *c. elegans*[98,99].

While ZFNs have been shown to function efficiently to cleave DNA at their intended target site, it remains an open question the extent of unintended off-target breaks that may be occurring throughout the genome. Initial attempts to identify the sites of off-target mutagenesis involved performing SELEX to determine the DNA-binding profiles of monomeric ZFNs and then using these preferences to identify closely matched sites in the genome[42]. Analyzing the top 15 most similar sites identified by this method in cells treated with ZFNs targeted to *CCR5*, scientists at Sangamo Biosciences found off-target cleavage at the highly related *CCR2* locus and very rare mutagenesis at a second site. The same method was used to analyze off-target breaks in clonal human embryonic stem cell lines that had been modified with ZFNs and in examining 10 potential off-target loci for five different ZFN pairs, identified only one instance of off-target mutagenesis[43]. The inherent flaw in this method, however, is that it examines binding site profiles of monomeric ZFNs and then attempts to extrapolate this to assess cleavage profiles of the dimeric nuclease. Recently, two groups developed improved methods to assay off-target cleavage induced by dimeric ZFNs. The von Kalle group took advantage of the tendency of integrase-deficient lentivirus (IDLV) to insert into DSBs in the genome to map breaks induced by ZFN treatment[100]. Using linear amplification-mediated (LAM) PCR of the integrated IDLV, they identified four novel off-targets of the *CCR5* ZFNs. Another method described by Liu and colleagues uses an *in vitro* cleavage assay to identify sites

from a large combinatorial library that are cleaved by ZFNs[101]. Of the sites identified directly in this *in vitro* assay for the same pair of *CCR5* ZFNs evaluated by the von Kalle group, 37 occur in the human genome and nine of these showed evidence of mutagenesis in cells treated with these nucleases. While both of these studies improved upon existing methods and identified novel off-targets of the *CCR5* ZFNs, the sets of sites identified by the two methods overlap only at the closely related *CCR2* site, clearly demonstrating that neither approach has exhaustively profiled the spectrum of off-targets. Full characterization of the spectrum of ZFN off-target effects remains a focus of ongoing effort in the field.

TALE Nucleases (TALENs)

Building off the foundation laid by a decade of ZFN-induced genome editing, the discovery of TALEs as a novel DNA-binding domain was quickly followed by the engineering of TALE-nucleases (TALENs). Like ZFNs, TALEs were fused to the catalytic domain of the FokI endonuclease and shown to function as dimers to cleave their intended DNA target site[48,49,53,102]. Optimization of TALENs involved variably truncating the C-terminal domain and investigating the role that this plays in the spacing between the TALE binding site and the location of the induced DSB[48,53]. Highly active TALENs were engineered with several different C-terminal domains, however, the majority of TALENs described in the literature to date use the Δ152 N-terminal and +63 C-terminal domains originally described by Miller et al[48] with a spacer of 16-18 base pairs between the TALE binding sites.

Similar to ZFNs, TALENs have been shown to efficiently induce both NHEJ and HDR in human somatic[48,53,69] and pluripotent stem cells[85,103] as well as in many whole organisms,

including plants[62], zebrafish[56,58,104,105], mice[106] and pigs[107]. One of the major benefits of TALENs is that, unlike ZFNs, they can be engineered to target virtually any DNA sequence. With the only targeting restriction being the requirement for a 5' T, which is specified by the N-terminal domain, it has been calculated that at least three TALEN pairs (of array and spacer lengths known to function efficiently) can be engineered for every base pair of random DNA sequence[69]. This is in stark contrast to the one in 200 base pair targeting range of OPEN ZFNs. This unlimited targeting range, in addition to the ease with which TALENs can be engineered, makes them an attractive platform for genome editing. Conversely, the large size and repetitive nature of TALE arrays is a major challenge of this platform. A coding sequence of a TALEN is over three times larger than that of a ZFN specifying the same size target site and this can prohibit expression of TALENs using viral systems with limited packaging capabilities. Additionally, the unstable nature of tandemly repeated sequences makes it challenging to package TALE repeat-encoding sequences in viral systems and it has been shown that while they can be packaged into adenovirus, TALENs delivered by lentivirus are highly susceptible to rearrangements[108].

As with ZFNs, the potential of TALENs to induce off-target DSBs in the genome remains largely unexplored. Parallel to their work with ZFNs, Jaenisch and colleagues used SELEX to profile the specificity of a pair of TALENs targeted to the AAVS1 integration site and found evidence of NHEJ at 2 of the top 19 sites in ES cells that had undergone TALEN-induced gene targeting[103]. Another study used exome sequencing and low-coverage whole genome sequencing to profile four ES cell clones that had undergone gene targeting with TALENs and found no evidence of indels at unintended sites[85]. Despite this encouraging lack of off-target

NHEJ, the use of clonal lines in both of these studies makes it virtually impossible to detect low-level mutagenesis. Additional, and more in-depth, studies will be required to fully examine the genome-wide specificity of engineered TALENs.

Meganucleases

Meganuclease technology involves re-engineering the DNA-binding specificity of naturally occurring homing endonucleases, such as I-*Cre*I. I-*Cre*I belongs the LAGLIDADG family of endonucleases and functions as a homodimer to induce DSBs. Through a combination of rational design and selection, this homing endonuclease has been reengineered to target sequences in the human *RAG1* and *XPC* genes[109-111]. Engineered meganucleases have also been used to induce gene targeting in maize, rats and mice[112,113] as well as to target the genome of herpes simplex virus in african monkey fibroblasts, resulting in a reduced viral load[114]. While these studies show promise for the use of meganucleases in genome editing, the DNA-binding and cleavage domains of homing endonucleases are difficult to separate, and the relative difficulty of engineering proteins with novel specificities makes this technology less attractive than ZFN and TALEN alternatives. One potential advantage of meganuclease technology is that the mechanism of DSB-induction results in a 3' overhang, which may be more recombinagenic for HDR than the 5' overhang generated from cleavage by FokI. Additionally, meganucleases are the smallest class of engineered nucleases, making them potentially amenable to all standard gene delivery methods. Very little is known about the genome-wide specificity of engineered meganucleases in living cells. Off-target cleavage in cells treated with two engineered meganucleases was assessed by quantifying γ-H2AX foci and was found to be

comparable to cells treated with I-SceI[110,111]. While meganucleases performed favorably in this very coarse measurement of DSBs, no studies have characterized the detailed binding profiles of engineered meganucleases and the extent to which these enzymes bind off-target sites in the genome remains to be defined.

CRISPR/Cas nucleases

The recent finding that the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system can be adapted to induce site-specific DSBs has propelled the CRISPR/Cas system into the spotlight of the genome editing field. The CRISPR/Cas system evolved in bacteria as a defense mechanism against invading plasmids and viruses and relies on uptake of foreign DNA fragments into CRISPR loci and subsequent transcription and processing into short CRISPR RNAs (crRNAs), which then anneal to trans-activating crRNA (tracrRNA) and direct Cas-mediated cleavage of complementary foreign nucleic acids[115,116]. Recent studies have shown that using a single crRNA-tracrRNA fusion guide RNA (gRNA), this modified type II CRISPR system can direct Cas9 to cleave novel DNA targets specified by 20 nucleotides in the 5' end of the gRNA[117,118]. Following their discovery, these RNA-Guided Nucleases (RGNs) were rapidly applied to genome editing in human somatic and pluripotent cells[119-121], bacteria[122], yeast[123], zebrafish[124] and mice[125]. Unlike the three systems discussed above, RGNs do not require the engineering of novel proteins for each DNA target site and virtually any sequence can be targeted with the only requirement being an NGG sequence, known as the protospacer adjacent motif (PAM), immediately 3' to the 20 nucleotide genomic target DNA site [117]. The relative ease with which new sites can be targeted, simply by altering

the 20 nucleotides of the gRNA that dictate specificity, makes this system a highly attractive method for introducing site-specific DSBs. Additionally, because the Cas9 protein is not directly coupled to the gRNA, which dictates sequence specificity, this system may be highly amenable to multiplexing through the concurrent use of multiple gRNAs to induce simultaneous DSBs at several loci. Very little is known about the specificity of RGNs, however, studies performed *in vitro* and in bacteria suggest that mutations in the target site, as well as non-canonical PAM sequences, are well-tolerated[117,122]. While the simplicity of the RGN system is appealing, this newcomer to the field is virtually untested compared to the years, or decades, of research that have been spent studying other engineered DNA-binding proteins, and it remains to be seen how robustly and specifically RGNs will function for genome editing.

Applications of Genome Editing

Genetic engineering has long been used to manipulate the mouse genome for the purpose of generating mouse models. This was made possible by the existence of mouse embryonic stem cells, which could be easily genetically manipulated in culture due to their high endogenous rates of homologous recombination, and subsequently implanted into pseudopregnant mice where they would contribute to the formation of chimeric embryos. This technology, however, was limited to mice and was not readily adaptable to other model organisms due to either a lack of embryonic stem cells or low rates of homologous recombination, which made genetic manipulation unfeasible. The study of many model organisms relied on imperfect technologies such as morpholino-mediated inhibition of target genes in zebrafish, which leads to incomplete knockdown and a large number of off-target

effects. The ability to use targeted nucleases to introduce site-specific mutations or stimulate HDR with donor DNA molecules has enabled gene editing in a number of organisms where it was not previously possible. Plants, zebrafish and rats are the most well-known beneficiaries of this technology[34,39,81,82,96] but it has also been applied to genetic manipulation of a wide variety of other model organisms such as sea urchins, crickets, and pigs[107,126,127].

Human embryonic stem (ES) cells and induced pluripotent stem (iPS) cells hold great promise as disease models and for potential gene therapy applications, however, unlike mouse ES cells, rates of HDR in human ES and iPS cells are low and require the use of large donor templates[128]. The use of nucleases to stimulate HDR has made it possible to genetically engineer human ES and iPS cells in a more robust fashion[31,43,88,103,119]. The ability to edit the genomes of these previously intractable cells opens the door to the possibilities of disease modeling and, potentially, to the development of gene therapies for the treatment of genetic disorders.

These technologies also make it possible to edit the genome of a wide variety of somatic cells. The ability to knockout target genes makes it possible to study the function of specific genes and gene products in the relevant cell type[80] and the efficiency of nuclease-induced NHEJ even makes it possible to knockout multiple genes in a single step[129]. Introduction of targeted point mutations should enable functional assessment of SNPs and disease modeling in relevant populations of somatic cells. Additionally, because nuclease-induced HDR can be used to introduce cassettes as large as 8kb, it is possible to insert entire transgenes into safe harbor loci or to tag endogenous genes[87]. Direct insertion into a defined locus allows for the study of different transgenes in an isogenic setting. The general applicability of this is illustrated by a

study in which various different glucocorticoid receptor reporter constructs were inserted into the AAVS1 integration site to generate a panel of isogenic cell lines in which the various reporters could be studied in identical chromatin contexts[130]. The same study also generated another panel of cell lines, each expressing a different shRNA targeted to varying components of the *mTOR* pathway. The ability to tag endogenous genes and to generate isogenic cell lines for the study of disease mutations or of regulation of reporter constructs eliminates many of the confounding variables that can accompany studies that use patient-derived human cells. Thus, transgenic human transformed and primary somatic cells are a powerful tool for studying the regulation and function of genes and for disease modeling in differentiated cell types.

Genome editing technology has had a dramatic impact on the field of biology by enabling the previously impossible manipulation of the genomes of almost any organism or cell type. The broad importance of this technology is evidenced by the fact that nuclease-induced genome editing was recognized by Nature Methods as the 2011 "Method of the Year" and by Science as runner-up for "Breakthrough of the Year" for 2012. Additionally, there is great potential for the use of this technology in therapeutic approaches to treat genetic disorders. Indeed, ZFNs targeted to the gene encoding CCR5, the major co-receptor used for HIV infection, are currently in phase 2 clinical trials for the treatment of HIV/AIDs. Genome editing technology is being applied to a rapidly expanding list of organisms and cell types, and many more applications of this broadly useful tool are surely not far behind.

Engineered Transcription Factors

Eukaryotic gene expression is controlled by a complex network of transcription factors. Unlike in prokaryotes, most promoters in eukaryotes exist in an inactive ground state and require the complex, and often combinatorial, actions of multiple transcriptional activators for gene expression[131-133]. Synthetic transcription factors have potentially broad applications in both research and therapeutics. The capability to specifically alter expression of endogenous target genes using these artificial proteins will permit in-depth studies of gene regulation and of the phenotypic effects resulting from altered expression of a given gene. Direct regulation of the endogenous gene has several advantages over alternative methods such as RNA interference (RNAi) or cDNA overexpression. While RNAi is a simple method for suppressing gene expression in mammalian cells[134], the myriad of off-target effects can confound interpretation of any observed phenotypic effects[135,136]. Introduction of an exogenous cDNA-encoded transcript results in robust overexpression of the gene. However, it can be difficult to control the level of expression as these are typically driven by strong constitutive promoters, and this method usually only allows for expression of a single transcript and ignores potential splice variants. The use of engineered transcription factors to modulate expression of the endogenous locus maintains alternative splicing, which can be important for full recapitulation of a phenotype[137]. Additionally, the level of expression can be titrated through the use of activator or repressor domains of varying strengths. These advantages make engineered transcription factors a powerful tool for studying the regulation and effects of specific gene targets and potentially for many therapeutic applications.

Zinc Finger Transcription Factors

Zinc finger-containing proteins represent one of the largest classes of naturally occurring transcription factors. In addition to their use as ZFNs, engineered zinc fingers have been fused to other effector domains to generate artificial transcription factors. The first engineered zinc finger transcription factor (ZTF) was targeted to a *BCR-ABL* fusion oncogene and was shown to activate transcription of a reporter construct when fused to the herpes simplex virus VP16 activation domain and to block transcription at an endogenous locus in the absence of an activation domain[16]. Similarly, a zinc finger targeted to the 5' UTR of the *erb-B2/HER2* locus was able to activate transcription of a luciferase reporter when fused to either the VP16 activation domain or a tetrameric repeat of VP16 (VP64) and was able to repress reporter gene expression when fused to either the Krupple-associated box (KRAB), ERF repressor domain (ERD) or the mSIN3 interaction domain (SID)[138]. Further studies demonstrated the ability of ZTFs to activate endogenous loci. ZF activators and repressors targeted to *ERB2* and *ERB3* were shown to regulate transcription of these endogenous genes in human, monkey and mouse cells[139]. Additionally, zinc fingers engineered to bind the promoter regions of the human vascular endothelial growth factor A (*VEGFA*) and erythropoietin (*EPO*) genes and fused to either VP16 or the activation domain of the p65 subunit of NF-κB were able to activate expression of the endogenous transcripts[140,141]. Further validation of the *VEGFA* ZF-activators demonstrated that they were able to induce VEGFA protein expression and stimulate angiogenesis when expressed in a mouse model[137]. The *VEGF-A* ZFs were also shown to function as repressors when fused to the ligand-binding domain of thyroid hormone receptor alpha (TRα), which functions as a transcriptional repressor in the absence of ligand, or to vErbA

which functions as a constitutive repressor[142]. Similarly, engineered zinc finger proteins fused to the KRAB repression domain and targeted to the HIV-1 promoter were able to repress HIV-1 viral replication[143]. ZF-activators have been shown to function in a variety of organisms, such as plants[144,145], *Drosophila*[146] and rats[147], and the therapeutic potential of this technology has been demonstrated by studies showing that ZF-activators targeted to the human γ-globin gene can induce fetal hemoglobin production in adult erythroblasts as well as in a transgenic mouse model[148,149].

In addition to using engineered ZTFs for targeted modulation of gene expression, combinatorial libraries of these proteins made by modular assembly can be screened for induction of a desired phenotype. Multiple groups have screened ZTF libraries for phenotypes such as enhanced expression of certain genes, drug resistance, thermotolerance or osmotolerance in yeast, and differentiation in mammalian cells[17,150-155]. This forward genetics approach enables the identification of transcription factors whose expression results in a desired phenotype, and then subsequent analysis of their targets to identify the genes and pathways involved in this process.

Specificity remains an issue for engineered ZTFs. Unlike dimeric ZFNs, which require two zinc finger arrays binding in a precise orientation in order to generate a functional enzyme, ZTFs function as monomers and are active at any site where the ZF is able to bind. *VEGFA* ZF-activators have recently failed to show efficacy in clinical trials as a treatment for diabetic neuropathy and it has been speculated that a lack of specificity of these proteins may have contributed to their failure[156,157]. This is not altogether surprising given that the ZF used in these trials was a three-finger protein and a 9bp binding site would be expected to occur over

20,000 times in the human genome. The 18bp binding site of a six-finger protein would theoretically be unique in the human genome, however, the strong affinity of three-finger proteins would likely allow subsets of the longer array to bind DNA and therefore decrease specificity. One way to increase the specificity of these longer proteins is to use three two-finger units separated by extended linkers which disrupt cooperative binding between the fingers[40]. In this way, it is possible to design six-finger proteins that bind their 18bp target site with high affinity and specificity as evidenced by a ZF-repressor targeted to *CHK2*, which was shown to only alter the expression of its target gene when expression of 16,000 genes were analyzed by microarray analysis[158]. An alternative approach to increasing specificity could be to split the effector and decrease the affinity of each zinc finger array monomer such that cooperative binding between the two is necessary to form a stable complex at the target site. If two weak affinity ZFs were fused, for example, to the NFκB p65 and p50 subunits, the favorable dimerization energy between the two subunits could serve to dramatically increase specific binding because cooperativity would only be in effect in the rare instances when the two ZF binding sites occur side-by-side.

TALE Transcription Factors

TALEs exist naturally as transcriptional activators that act as virulence factors through their ability to mimic eukaryotic transcription factors in plant host cells[159,160]. It was shown that using the native AvrHah1 TALE activation domain, but altering the repeats to confer novel specificity, an engineered TALE-activator could induce expression of the Bs3 gene in pepper plants[62]. Many groups subsequently adapted this platform for the regulation of gene

expression in mammalian cells by replacing the endogenous TALE activation domain with either VP16 or VP64[48,65,68,161-165]. The most robust activation was observed with a TALE-activator targeted to NTF3, which showed a 30-fold induction of gene expression in HEK293 cells[48]. Additionally, TALE-activators have been shown to function on endogenous genes in several other human cell lines [65,68,161-164] as well as in mouse embryonic stem cells[165]. Two studies have explored the use of engineered TALE-repressors; while a TALE fused to the EAR repression domain SRDX was able to repress the *RD29A* gene by 7-fold in *Arabidopsis*[166], testing of 8 different repression domains in human cells yielded a maximum repression of *SOX2* with a TALE-SID fusion of only 2.8-fold and further testing at the *CACNA1C* locus yielded repression ranging from 2 to 4-fold[164].

As with TALENs, little is known about the specificity of TALE-TFs. No study has explored the genome-wide expression profile of cells treated with these fusion proteins and so it remains to be seen how specific their effects are. Additionally, many of the TALE-TFs described in the literature induced only modest increases or decreases in target gene expression. While these results suggest that TALE-TFs may not function robustly to modulate expression of endogenous genes, the use of varied TALE architectures (for example, the number of TALE repeats, the length and sequence of the N- and C-terminal domains and the transcriptional activation and repression domains used) in these studies makes it difficult to ascertain which parameters may be influencing activity and to assess the true efficacy of engineered TALE-TFs. As with TALENs, however, the ease with which TALE arrays can be engineered makes TALE-TFs an attractive alternative to zinc finger-based transcription factors.

Engineered Proteins for Targeted Histone and DNA Modifications

Histone Modifications and DNA Methylation

Gene expression is controlled by superimposed layers of regulation, which begin with the specific promoter DNA sequence and go all the way up to higher order chromatin structures. The DNA sequence contains binding sites for transcription factors that directly control gene expression levels, as discussed in the previous section. Direct chemical modification of the DNA through cytosine methylation provides another level of regulation as methylation can either promote or prohibit transcription factor binding[167-169]. An additional layer of regulation derives from the packaging of DNA around nucleosomes, which comprise 147 base pairs of DNA wrapped around a histone octamer. Nucleosome positioning is critical for gene expression and is regulated by a combination of DNA sequence, nucleosome remodeling enzymes and DNA-binding proteins[170,171]. Histones are subject to a variety of post-translational modifications, such as methylation and acetylation, and these modifications contribute to the control of gene expression by recruiting protein complexes and influencing chromatin compaction[172,173].

DNA methylation at the 5' position of cytosines (5mC) is known to play critical roles in determining gene expression levels in development and disease. Regions of open chromatin are associated with hypomethylated DNA and actively transcribed genes, whereas hypermethylated DNA is packaged into inactive chromatin. In addition to dictating the tissue-specific gene expression and the developmental pathways of different cells, aberrant DNA methylation is highly linked to the initiation and progression of most cancers and has also been

associated with a wide variety of other human diseases such as lupus, diabetes and rheumatoid arthritis [174-176].

DNA methylation is catalyzed by DNA methyltransferases (DNMTs) and occurs primarily at CpG dinucleotides. DNMT1 shows a preference for hemi-methylated DNA and is responsible for maintenance of methylation after DNA replication while DNMT3a and DNMT3b are responsible for *de novo* methylation. Previously, the only known pathway for reversion of 5mC to cytosine was passive demethylation occurring during replication. However, a number of papers in the last few years have identified TET1 as an enzyme that catalyzes the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) and it has been suggested that this may induce active demethylation via deamination of 5hmC and subsequent base excision repair[177-180]. Additionally, it has been shown that the TET proteins actively promote demethylation by further oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which are then reverted to cytosine through base excision repair[181,182]. While the relative contributions of these two pathways to DNA demethylation in the cell remain speculative, it is widely accepted that TET proteins are responsible for the oxidation of 5mC to 5hmC and that the end result of this process is demethylation.

Histones undergo a large number of post-translational modifications, which can have either activating or repressive effects on gene transcription. Some of the most well understood marks are histone acetylation and tri-methylation of histone H3 lysine 4 (H3K4me3), which are associated with active promoters, tri-methylation of histone H3 lysine 36 (H3K36me3), which is found in actively transcribed regions, and tri-methylation of histone H3 lysines 9 and 27 (H3K9me3 and H3K27me3), both of which are associated with repressed genes[172,183-187].

Addition and removal of acetyl groups is catalyzed by histone acetyl transferase (HAT) complexes and by histone deacetylases (HDACs), respectively. Histone methyltransferases (HMTs) catalyze the addition of methyl groups from S-adenosylmethionine to histones; however, the exact mechanisms responsible for recruiting HMTs to specific targets remain an area of ongoing research. It is known that specific DNA sequences, known as Polycomb response elements (PREs), and trans-acting factors that recognize these PREs are responsible for recruitment of Polycomb repressive complex 2 (PRC2), which methylates H3K27, in *Drosophila*[188]. Similarly, transcriptionally inactive GC-rich regions have been shown to recruit PRC2 in human ES cells[189]. Additionally, it has been suggested that long non-coding RNAs (lncRNAs) may play a role in targeting methyltransferases and demethylases to specific genomic locations and this theory is supported by the finding that many lncRNAs, including *HOTAIR* and *RepA/Xist*, bind PRC2[190-193].

Targeted Histone Modification

While histone modifications are known to play many important roles in the regulation of gene expression, many of the exact mechanisms through which this occurs remain elusive. In many cases it is difficult to definitively discern whether histone modifications are causative in the activation or repression of target genes or whether they are a byproduct of whatever driving force is behind this process. Overexpressing or inhibiting enzymes such as histone methyltransferaces (HMTs) or histone deacetylases (HDACs) can cause broad changes in chromatin in a cell and give some insight into the functions of these modifications. However,

the global effect of these approaches makes it impossible to specifically examine the effects of histone modifications in a locus-specific manner.

To investigate the possibility of targeting histone modifications to a specific locus, Pabo and colleagues used an engineered ZF to direct histone methyltransferase activity[194]. A zinc finger targeted to the *VEGFA* locus was fused to the catalytic domains of either G9A or SUV39H1 and expression of these fusions in human cells resulted in a two to three-fold reduction in *VEGFA* expression. ChIP-seq revealed significantly increased levels of H3K9 methylation at the target locus. Interestingly, increased H3K9 methylation was observed as far as 900 base pairs downstream of the ZF binding site and the authors speculate that the ZF-induced methylation might act as a nucleation point for recruitment of HP1, which is known to bind methylated H3 through its chromodomain[195], and subsequent spreading of the methylation effect.

This single example demonstrates the feasibility of targeted histone modifiers for specifically manipulating chromatin status. Both the increase in H3K9 methylation and decrease in gene expression were relatively modest in this case, although simultaneous recruitment of HDAC activity did increase transcriptional repression. It is possible that co-recruitment of enzymes that catalyze different histone marks may serve as a general strategy for eliciting more robust changes in chromatin status and gene expression.

Targeted DNA Methylation

The ability to specifically target DNA methylation has great potential in therapeutics and as a basic research tool. Directing DNA methyltransferase (DNMT) activity would allow for,

potentially heritable, silencing of target loci and would enable researchers to explore the link between DNA methylation and gene expression. Initial attempts to engineer site-specific DNMTs focused on fusing zinc finger domains to bacterial methyltransferases such as M.*Sss*I, M.*Hha*I and M.*Hpa*II[196-198]. While these enzymes were able to increase methylation of cytosines near ZF binding sites in bacteria and yeast, they retained endogenous methyltransferase activity and therefore induced high levels of methylation throughout the genome. Attempts to improve the specificity of ZF-DNMT fusions included rational mutagenesis to reduce the endogenous binding activity of the methyltransferase domain[199] and utilization of a split methyltransferase enzyme[200]. While the mutant methyltransferase did bias methylation towards the intended target site, the inherent problem remained that binding of the ZF to its cognate site was not required for activity of the enzyme. Further studies revealed that activity of the split methyltransferase was not reliant on zinc finger binding and demonstrated that the protein methylated target and non-target sites with equal efficiency[201]. Attempts to improve on this approach through the use of a naturally heterodimeric methyltransferase enzyme met with some success, however, off-target methylation was still observed at higher concentrations of enzyme[202]. More recent experiments have focused on fusions to the catalytic domain of the *de novo* mammalian methyltransferases DNMT3a and DNMT3b[203,204]. These enzymes were shown to direct methylation of target sites in reporters and endogenous genes in human cells, however, as with previous studies, off-target methylation was observed. Additionally, the rate of methylation at on-target sites, and consequent repression of target genes, was low. An attempt to improve activity by engineering a single-chain fusion of the catalytic domains of DNMT3a and DNMT3L, which has been shown to stimulate the activity of DNMT3a, was able to

increase methylation activity, but only by 2-fold[205]. Overall, considerable progress has been made in engineering site-specific DNA methyltransferases, but there remains significant room for improvement, specifically in enhancing the rates of on-target methylation and in abolishing off-target methylase activity.

Introduction to the Thesis

This dissertation uses two of the most commonly engineered DNA-binding domains, zinc fingers (ZFs) and transcription activator-like effector (TALE) repeats, to target the function of nucleases, transcriptional activators and a DNA hydroxymethylase.

Chapter 2: "In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases"

At the time this study was undertaken, although it was well established that ZFNs could enable robust gene targeting in human pluripotent stem cells, there had been no demonstration that this method could be used to correct a disease-causing mutation. Despite the fact that nuclease-induced genome editing was often proposed as a highly promising approach for gene therapy, the critical proof-of-principle showing correction of a disease gene had yet to be performed. We therefore undertook this project to demonstrate ZFN-mediated correction of the genetic mutation responsible for sickle cell anemia in human induced pluripotent stem cells (hiPSCs).

At the time this study was initiated, ZFNs were the state of the art engineered nucleases. Prior to the discovery of the TALE DNA recognition code, ZFNs and meganucleases

were the only platforms available for inducing site-specific DSBs. Our lab had recently developed OPEN, a robust method for engineering novel zinc fingers with high affinities and specificities for their target sites. We took advantage of this technology to engineer two pairs of ZFNs targeted to the human β-globin locus and demonstrated both were able to cleave their intended target site in human somatic cells. In collaboration with Marius Wernig's laboratory at Stanford University, we used ZFN-induced gene targeting to correct the sickle cell mutation in iPS cell lines derived from two sickle cell anemia patients.

Several other studies published at the same time underscored the importance and broad applicability of this work. Cheng and colleagues conducted a parallel study in which they also demonstrated successful ZFN-induced genetic correction of the sickle cell mutation in patient iPS cells[206]. Similarly, Jaenisch and colleagues used ZFNs to correct a Parkinson's disease-associated mutation in the α-synuclein gene of patient iPS cells in order to generate a pair of isogenic cell lines to enable study of the disease[94]. Yet another study used a combination of ZFNs and *piggybac* technology to correct a mutation responsible for α_1-antitrypsin deficiency in patient-derived iPS cells[207]. Together with these concurrently published studies, the work described in Chapter 2 of this dissertation demonstrates an important proof-of-principle for the use of engineered nucleases in the correction of genetic disorders.

Chapter 3: "Robust, synergistic regulation of human gene expression using TALE activators"

In the same way that the increased targeting range and ease of engineering of TALENs have made them attractive alternatives to ZFNs, TALE-transcription factors (TALE-TFs) provided

an easier-to-use alternative to zinc finger transcription factors (ZTFs). Additionally, we were interested in the use of TALEs for generating other monomeric fusion proteins, as has been described for ZFs (e.g.--ZF-methyltransferases). However, the published literature painted a bleak picture of the activities and efficiencies of monomeric TALE fusion proteins. While dimeric TALENs had been shown to efficiently cleave their target sites with high rates of activity, monomeric TALE fusion proteins described in the literature displayed disappointingly low activities and success rates. One TALE-activator, constructed using the Δ152 N-terminal and +95 C-terminal domains and fused to the VP16 transcriptional activator had been shown to induce expression of the human *NTF3* gene by 30-fold[48]. However, of the other 25 other TALE-activators described in the literature, for which quantitative information was available, only three were able to induce target gene expression by five-fold or more[62,65,161-163,165]. Examination of these studies revealed that the described TALE-TFs were extremely varied in their TALE architectures and in the effector domains used. Additionally, experiments conducted across a variety of organisms and cell types made it difficult to draw definitive conclusions as to the true success rates of these proteins. We therefore undertook a large-scale assessment of TALE-activator activity, on a single, common framework, in order to determine the robustness of this technology.

Using as a starting point the single robust TALE-activator described in the literature, we optimized the number of TALE repeats required for maximal activation of an endogenous gene. Using this robust architecture, we then explored the broad applicability of this technology by engineering TALE-activators capable of efficiently activating an additional protein-coding gene, as well as a microRNA cluster. Additionally, we found that concurrent expression of multiple

TALE-activators targeted to a single locus resulted in synergistic increases in target gene expression.

The work described here changed the outlook for TALE-activators as a robust method for targeted gene activation. Contrary to the published literature at the time, we demonstrate that TALE-activators can robustly induce gene transcription with high success rates. A parallel study published in the same issue of Nature Methods similarly showed high success rates of engineered TALE-activators and confirmed the ability of multiple proteins to activate gene expression synergistically [208]. In addition to establishing TALE-activators as a powerful technology for the manipulation of gene expression, our work demonstrates the feasibility of engineering robust monomeric TALE fusion proteins. This raises the possibility of using TALEs to confer sequence specificity to a myriad of other enzymes, such as those that catalyze changes in DNA methylation or in the covalent modification of histones.

Chapter 4: "Targeted DNA demethylation and gene activation using engineered TALE-TET proteins"

Targeted DNA methylation has been achieved by fusing ZFs to DNA methyltransferases, however, no study has yet explored the targeted removal of cytosine methylation marks. Recent work has identified TET1 as a DNA hydroxymethylase, responsible for catalyzing conversion of 5mC to 5hmC and subsequent reversion to unmethylated cytosine. Encouraged by our successful engineering of robust monomeric TALE-activators, we sought to explore the potential to create a site-specific demethylase by fusing TET1 to engineered TALEs and ZFs.

Initial optimization of TALE-TET1 fusion proteins revealed that linkage of the catalytic domain of TET1 to the C-terminus of engineered TALEs induced robust demethylation of nearby CpGs. ZF fusions to this catalytic domain were similarly able to direct targeted demethylase activity. Implementation of this technology at two additional endogenous loci in human cells demonstrated efficient demethylation and, importantly, also showed a corresponding increase in target gene expression.

This work describes the first implementation of TALEs to direct sequence specificity of an enzyme that catalyzes changes in DNA methylation status. Furthermore, this is the first demonstration of site-specific demethylation. This technology holds great promise as a basic research tool for exploring causal links between DNA methylation and gene expression. Prior to this work, the ability to remove methylation marks was limited to methods which effect change on a global level. Chemical modulators, such as 5-azacytidine, efficiently demethylate DNA, however the genome-wide effect of this treatment makes it difficult to link phenotypes to specific methylation events. The ability to specifically demethylate target CpGs will allow researchers to dissect apart the effects of individual methylation marks and determine causality in the relationship between methylation and gene expression. In the longer-term, our work also raises the possibility of using engineered ZF or TALE technology to direct the targeted demethylation of specific gene sequences for potential therapeutic applications.

References

1. Tupler, R., Perini, G. & Green, M. R. Expressing the human genome. *Nature* **409,** 832–833 (2001).
2. J Miller, A. D. M. A. K. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO Journal* **4,** 1609 (1985).
3. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245,** 635–637 (1989).
4. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252,** 809–817 (1991).
5. Elrod-Erickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure* **4,** 1171–1180 (1996).
6. Desjarlais, J. R. & Berg, J. M. Redesigning the DNA-binding specificity of a zinc finger protein: a data base-guided approach. *Proteins* **12,** 101–104 (1992).
7. Desjarlais, J. R. & Berg, J. M. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc Natl Acad Sci USA* **89,** 7345–7349 (1992).
8. Pavletich, N. P. & Pabo, C. O. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261,** 1701–1707 (1993).
9. Elrod-Erickson, M., Benson, T. E. & Pabo, C. O. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6,** 451–464 (1998).
10. Pabo, C. O. & Nekludova, L. Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *Journal of Molecular Biology* **301,** 597–624 (2000).
11. Rebar, E. J. & Pabo, C. O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263,** 671–673 (1994).
12. Choo, Y. & Klug, A. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci USA* **91,** 11163–11167 (1994).
13. Jamieson, A. C., Kim, S. H. & Wells, J. A. In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33,** 5689–5695 (1994).
14. Wu, H., Yang, W. P. & Barbas, C. F. Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci USA* **92,** 344–348 (1995).
15. Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci USA* **97,** 7382–7387 (2000).
16. Choo, Y., Sánchez-García, I. & Klug, A. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372,** 642–645 (1994).
17. Bae, K.-H. *et al.* Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21,** 275–280 (2003).
18. Segal, D. J., Dreier, B., Beerli, R. R. & Barbas, C. F. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5‘-GNN-3’ DNA

target sequences. *Proc Natl Acad Sci USA* **96,** 2758–2763 (1999).
19. Dreier, B. Development of Zinc Finger Domains for Recognition of the 5'-ANN-3' Family of DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. *Journal of Biological Chemistry* **276,** 29466–29478 (2001).
20. Dreier, B. Development of Zinc Finger Domains for Recognition of the 5'-CNN-3' Family DNA Sequences and Their Use in the Construction of Artificial Transcription Factors. *Journal of Biological Chemistry* **280,** 35588–35597 (2005).
21. Liu, Q., Xia, Z., Zhong, X. & Case, C. C. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277,** 3850–3856 (2002).
22. Wolfe, S. A., Greisman, H. A., Ramm, E. I. & Pabo, C. O. Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *Journal of Molecular Biology* **285,** 1917–1934 (1999).
23. Isalan, M., Choo, Y. & Klug, A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci USA* **94,** 5617–5621 (1997).
24. Isalan, M., Klug, A. & Choo, Y. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37,** 12026–12033 (1998).
25. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5,** 374–375 (2008).
26. Kim, H. J., Lee, H. J., Kim, H., Cho, S. W. & Kim, J.-S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* **19,** 1279–1288 (2009).
27. Greisman, H. A. A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. *Science* **275,** 657–661 (1997).
28. Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O. & Joung, J. K. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci USA* **100,** 12271–12276 (2003).
29. Cornu, T. I. *et al.* DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* **16,** 352–358 (2008).
30. Maeder, M. L. *et al.* Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* **31,** 294–301 (2008).
31. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5,** 97–110 (2009).
32. Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K. & Porteus, M. H. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther* **16,** 707–717 (2008).
33. Söllü, C. *et al.* Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res* **38,** 8269–8276 (2010).
34. Townsend, J. A. *et al.* High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459,** 442–445 (2009).
35. Zhang, F. *et al.* High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proc Natl Acad Sci USA* **107,** 12028–12033 (2010).
36. Foley, J. E. *et al.* Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS ONE* **4,** e4348 (2009).
37. Sander, J. D. *et al.* Selection-free zinc-finger-nuclease engineering by context-

dependent assembly (CoDA). *Nat Methods* **8,** 67–69 (2011).
38. Isalan, M., Klug, A. & Choo, Y. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter : Article : Nature Biotechnology. *Nat Biotechnol* **19,** 656–660 (2001).
39. Doyon, Y. *et al.* Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26,** 702–708 (2008).
40. Moore, M., Klug, A. & Choo, Y. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci USA* **98,** 1437–1441 (2001).
41. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435,** 646–651 (2005).
42. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26,** 808–816 (2008).
43. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* **27,** 851–857 (2009).
44. Pearson, H. Protein engineering: The fate of fingers. *Nature* **455,** 160–164 (2008).
45. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326,** 1509–1512 (2009).
46. Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326,** 1501–1501 (2009).
47. Morbitzer, R., Römer, P., Boch, J. & Lahaye, T. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc Natl Acad Sci USA* **107,** 21617–21622 (2010).
48. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29,** 143–148 (2010).
49. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186,** 757–761 (2010).
50. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* **39,** 359–372 (2011).
51. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335,** 720–723 (2012).
52. Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335,** 716–719 (2012).
53. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39,** 9283–9293 (2011).
54. Zhang, H. S. *et al.* A Designed Zinc-finger Transcriptional Repressor of Phospholamban Improves Function of the Failing Heart. *Mol Ther* **20,** 1508–1515 (2012).
55. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* **22,** 1716–1720 (2012).
56. Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* – (2012). doi:10.1038/nature11537
57. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14,** 49–55 (2013).
58. Sander, J. D. *et al.* Targeted gene disruption in somatic zebrafish cells using engineered

TALENs. *Nat Biotechnol* **29,** 697–698 (2011).
59. Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* **29,** 699–700 (2011).
60. Reyon, D., Khayter, C., Regan, M. R., Joung, J. K. & Sander, J. D. Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Curr Protoc Mol Biol* **Chapter 12,** Unit 12.15 (2012).
61. Morbitzer, R., Elsaesser, J., Hausner, J. & Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res* **39,** 5790–5799 (2011).
62. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39,** e82 (2011).
63. Li, L. *et al.* Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol Biol* **78,** 407–416 (2012).
64. Weber, E., Gruetzner, R., Werner, S., Engler, C. & Marillonnet, S. Assembly of designer TAL effectors by Golden Gate cloning. *PLoS ONE* **6,** e19722 (2011).
65. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29,** 149–153 (2011).
66. Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7,** 171–192 (2012).
67. Briggs, A. W. *et al.* Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res* (2012). doi:10.1093/nar/gks624
68. Wang, Z. *et al.* An integrated chip for the high-throughput synthesis of transcription activator-like effectors. *Angew. Chem. Int. Ed. Engl.* **51,** 8505–8508 (2012).
69. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* (2012). doi:10.1038/nbt.2170
70. Kim, Y. *et al.* A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* **31,** 251–258 (2013).
71. Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Höning, K. & Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes. *Nat Biotechnol* **31,** 76–81 (2012).
72. Friedmann, T. & Roblin, R. Gene therapy for human genetic disease? *Science* **175,** 949–955 (1972).
73. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci USA* **91,** 6064–6068 (1994).
74. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14,** 8096–8106 (1994).
75. Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res* **23,** 5012–5019 (1995).
76. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol* **15,** 1968–1973 (1995).

77. Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO Journal* **17,** 5497–5508 (1998).
78. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4,** 712–720 (2003).
79. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161,** 1169–1175 (2002).
80. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA* **105,** 5809–5814 (2008).
81. Geurts, A. M. *et al.* Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325,** 433 (2009).
82. Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26,** 695–701 (2008).
83. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33,** 25–35 (1983).
84. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* **8,** 753–755 (2011).
85. Ding, Q. *et al.* A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* **12,** 238–251 (2013).
86. Bibikova, M. *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21,** 289–297 (2001).
87. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci USA* **104,** 3055–3060 (2007).
88. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25,** 1298–1306 (2007).
89. Li, L., Wu, L. P. & Chandrasegaran, S. Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci USA* **89,** 4275–4279 (1992).
90. Kim, Y. G. & Chandrasegaran, S. Chimeric restriction endonuclease. *Proc Natl Acad Sci USA* **91,** 883–887 (1994).
91. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* **93,** 1156–1160 (1996).
92. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300,** 763 (2003).
93. Sebastiano, V. *et al.* In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells* **29,** 1717–1726 (2011).
94. Soldner, F. *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146,** 318–331 (2011).
95. Cai, C. Q. *et al.* Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol Biol* **69,** 699–709 (2009).
96. Shukla, V. K. *et al.* Precise genome modification in the crop species Zea mays using zinc-

finger nucleases. *Nature* **459,** 437–441 (2009).
97. Meyer, M., de Angelis, M. H., Wurst, W. & Kühn, R. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc Natl Acad Sci USA* **107,** 15022–15026 (2010).
98. Morton, J., Davis, M. W., Jorgensen, E. M. & Carroll, D. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. *Proc Natl Acad Sci USA* **103,** 16370–16375 (2006).
99. Wood, A. J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333,** 307 (2011).
100. Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **29,** 816–823 (2011).
101. Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* **8,** 765–770 (2011).
102. Li, T. *et al.* Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* **39,** 6315–6325 (2011).
103. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29,** 731–734 (2011).
104. Cade, L. *et al.* Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res* (2012). doi:10.1093/nar/gks518
105. Zu, Y. *et al.* TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods* **10,** 329–331 (2013).
106. Sung, Y. H. *et al.* Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol* **31,** 23–24 (2013).
107. Carlson, D. F. *et al.* Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci USA* **109,** 17382–17387 (2012).
108. Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res* **41,** e63 (2013).
109. Arnould, S. *et al.* Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *Journal of Molecular Biology* **371,** 49–65 (2007).
110. Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456,** 107–111 (2008).
111. Grizot, S. *et al.* Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res* **37,** 5405–5419 (2009).
112. Gao, H. *et al.* Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J.* **61,** 176–187 (2010).
113. Ménoret, S. *et al.* Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. *FASEB J.* **27,** 703–711 (2013).
114. Grosse, S. *et al.* Meganuclease-mediated Inhibition of HSV1 Infection in Cultured Cells. *Mol Ther* **19,** 694–702 (2011).
115. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–338 (2012).
116. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea.

Science **327,** 167–170 (2010).
117. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
118. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109,** E2579–86 (2012).
119. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
120. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
121. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31,** 230–232 (2013).
122. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31,** 233–239 (2013).
123. DiCarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res* (2013). doi:10.1093/nar/gkt135
124. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31,** 227–229 (2013).
125. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* (2013). doi:10.1038/cr.2013.46
126. Ochiai, H. *et al.* Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes Cells* **15,** 875–885 (2010).
127. Watanabe, T. *et al.* Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat Commun* **3,** 1017 (2012).
128. Zwaka, T. P. & Thomson, J. A. *Homologous recombination in human embryonic stem cells. Nat Biotechnol* **21,** 319–321 (2003).
129. Liu, P.-Q. *et al.* Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol. Bioeng.* **106,** 97–105 (2010).
130. Dekelver, R. C. *et al.* Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res* **20,** 1133–1142 (2010).
131. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386,** 569–577 (1997).
132. Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98,** 1–4 (1999).
133. Ptashne, M. Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci.* **30,** 275–279 (2005).
134. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806–811 (1998).
135. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21,** 635–637 (2003).
136. Fedorov, Y. *et al.* Off-target effects by siRNA can induce toxic phenotype. *RNA* **12,** 1188–1196 (2006).
137. Rebar, E. J. *et al.* Induction of angiogenesis in a mouse model using engineered

transcription factors. *Nat. Med.* **8,** 1427–1432 (2002).
138. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci USA* **95,** 14628–14633 (1998).
139. Beerli, R. R., Dreier, B. & Barbas, C. F. Positive and negative regulation of endogenous genes by designed transcription factors. *Proc Natl Acad Sci USA* **97,** 1495–1500 (2000).
140. Liu, P.-Q. *et al.* Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem* **276,** 11323–11334 (2001).
141. Zhang, L. *et al.* Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *J Biol Chem* **275,** 33850–33860 (2000).
142. Snowden, A. W. *et al.* Repression of vascular endothelial growth factor A in glioblastoma cells using engineered zinc finger transcription factors. *Cancer Res.* **63,** 8968–8976 (2003).
143. Segal, D. J. *et al.* Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. *J Biol Chem* **279,** 14509–14519 (2004).
144. Guan, X. *et al.* Heritable endogenous gene regulation in plants with designed polydactyl zinc finger transcription factors. *Proc Natl Acad Sci USA* **99,** 13296–13301 (2002).
145. Holmes-Davis, R. *et al.* Gene regulation in planta by plant-derived engineered zinc finger protein transcription factors. *Plant Mol Biol* **57,** 411–423 (2005).
146. Jamieson, A. C. *et al.* Controlling gene expression in Drosophila using engineered zinc finger protein transcription factors. *Biochem Biophys Res Commun* **348,** 873–879 (2006).
147. Laganière, J. *et al.* An Engineered Zinc Finger Protein Activator of the Endogenous Glial Cell Line-Derived Neurotrophic Factor Gene Provides Functional Neuroprotection in a Rat Model of Parkinson's Disease. *jneurosci.org.ezp-prod1.hul.harvard.edu*
148. Wilber, A. *et al.* A zinc-finger transcriptional activator designed to interact with the gamma-globin gene promoters enhances fetal hemoglobin production in primary human adult erythroblasts. *Blood* **115,** 3033–3041 (2010).
149. Costa, F. C. *et al.* Induction of Fetal Hemoglobin In Vivo Mediated by a Synthetic γ-Globin Zinc Finger Activator. *Anemia* **2012,** 1–8 (2012).
150. Blancafort, P., Magnenat, L. & Barbas, C. F. Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol* **21,** 269–274 (2003).
151. Park, K.-S. *et al.* Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol* **21,** 1208–1214 (2003).
152. Lee, D.-K., Kim, Y. H., Kim, J.-S. & Seol, W. Induction and characterization of taxol-resistance phenotypes with a transiently expressed artificial transcriptional activator library. *Nucleic Acids Res* **32,** e116 (2004).
153. Blancafort, P. *et al.* Genetic reprogramming of tumor cells by zinc finger transcription factors. *Proc Natl Acad Sci USA* **102,** 11716–11721 (2005).
154. Blancafort, P. *et al.* Modulation of drug resistance by artificial transcription factors. *Mol. Cancer Ther.* **7,** 688–697 (2008).

155. Lee, J. *et al.* Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library. *PLoS ONE* **6,** e21112 (2011).
156. Rebar, E. J. Development of pro-angiogenic engineered transcription factors for the treatment of cardiovascular disease. *Expert Opin Investig Drugs* **13,** 829–839 (2004).
157. Eisenstein, M. Sangamo's lead zinc-finger therapy flops in diabetic neuropathy. *Nat Biotechnol* **30,** 121–123 (2012).
158. Tan, S. *et al.* Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci USA* **100,** 11997–12002 (2003).
159. Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* **318,** 648–651 (2007).
160. Römer, P. *et al.* Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318,** 645–648 (2007).
161. Geissler, R. *et al.* Transcriptional activators of human genes with programmable DNA-specificity. *PLoS ONE* **6,** e19509 (2011).
162. Tremblay, J. P., Chapdelaine, P., Coulombe, Z. & Rousseau, J. Transcription Activator-Like Effector Proteins Induce the Expression of the Frataxin Gene. *Hum Gene Ther* (2012). doi:10.1089/hum.2012.034
163. Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* **40,** 7584–7595 (2012).
164. Cong, L., Zhou, R., Kuo, Y.-C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3,** 968 (2012).
165. Bultmann, S. *et al.* Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* (2012). doi:10.1093/nar/gks199
166. Mahfouz, M. M. *et al.* Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol* **78,** 311–321 (2012).
167. Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* **21,** 4886–4892 (1993).
168. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405,** 486–489 (2000).
169. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16,** 6–21 (2002).
170. Wang, X., Bai, L., Bryant, G. O. & Ptashne, M. Nucleosomes and the accessibility problem. *Trends Genet.* **27,** 487–492 (2011).
171. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20,** 267–273 (2013).
172. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12,** 7–18 (2011).
173. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128,** 669–681 (2007).
174. Javierre, B. M. *et al.* Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* **20,** 170–179 (2010).
175. Rakyan, V. K. *et al.* Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet* **7,** e1002300 (2011).

176. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* **31,** 142–147 (2013).
177. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
178. Guo, J. U., Su, Y., Zhong, C., Ming, G.-L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).
179. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324,** 929–930 (2009).
180. Hajkova, P. *et al.* Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. *Science* **329,** 78–82 (2010).
181. He, Y. F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **333,** 1303–1307 (2011).
182. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333,** 1300–1303 (2011).
183. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120,** 169–181 (2005).
184. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436,** 876–880 (2005).
185. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
186. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41,** 376–381 (2009).
187. Kurdistani, S. K., Tavazoie, S. & Grunstein, M. Mapping global histone acetylation patterns to gene expression. *Cell* **117,** 721–733 (2004).
188. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10,** 697–708 (2009).
189. Mendenhall, E. M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6,** e1001244 (2010).
190. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464,** 1071–1076 (2010).
191. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329,** 689–693 (2010).
192. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322,** 750–756 (2008).
193. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40,** 939–953 (2010).
194. Snowden, A. W., Gregory, P. D., Case, C. C. & Pabo, C. O. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Curr Biol* **12,** 2159–2166 (2002).
195. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).
196. Xu, G. L. & Bestor, T. H. Cytosine methylation targetted to pre-determined sequences. *Nat Genet* **17,** 376–378 (1997).
197. McNamara, A. R., Hurd, P. J., Smith, A. E. F. & Ford, K. G. Characterisation of site-biased

DNA methyltransferases: specificity, affinity and subsite relationships. *Nucleic Acids Res* **30,** 3818–3830 (2002).
198. Carvin, C. D., Parr, R. D. & Kladde, M. P. Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins. *Nucleic Acids Res* **31,** 6493–6501 (2003).
199. Smith, A. E. & Ford, K. G. Specific targeting of cytosine methylation to DNA sequences in vivo. *Nucleic Acids Res* **35,** 740–754 (2007).
200. Nomura, W. & Barbas, C. F. In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase. *J. Am. Chem. Soc.* **129,** 8676–8677 (2007).
201. Meister, G. E., Chandrasegaran, S. & Ostermeier, M. An engineered split M.HhaI-zinc finger fusion lacks the intended methyltransferase specificity. *Biochem Biophys Res Commun* **377,** 226–230 (2008).
202. Meister, G. E., Chandrasegaran, S. & Ostermeier, M. Heterodimeric DNA methyltransferases as a platform for creating designer zinc finger methyltransferases for targeted DNA methylation in cells. *Nucleic Acids Res* **38,** 1749–1759 (2010).
203. Li, F. *et al.* Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. *Nucleic Acids Res* **35,** 100–112 (2007).
204. Rivenbark, A. G. *et al.* Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7,** 350–360 (2012).
205. Siddique, A. N. *et al.* Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. *Journal of Molecular Biology* **425,** 479–491 (2013).
206. Zou, J., Mali, P., Huang, X., Dowey, S. N. & Cheng, L. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* **118,** 4599–4608 (2011).
207. Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478,** 391–394 (2011).
208. Perez-Pinera, P. *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods* **10,** 239–242 (2013).

Chapter 2

In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases

Vittorio Sebastiano,[1,2,*] Morgan L. Maeder,[3,4,*] James F. Angstman,[3] Bahareh Haddad,[1] Cyd Khayter,[3] Dana T. Yeo,[1] Mathew J. Goodwin,[3] John S. Hawkins,[1] Cherie L. Ramirez,[3,4] Luis F.Z. Batista,[5] Steven E. Artandi,[5] Marius Wernig,[1,#] & J. Keith Joung[3,4,6,#]

[1]Institute for Stem Cell Biology & Regenerative Medicine, Department of Pathology, Stanford University School of Medicine, Stanford CA 94305 USA

[2]Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford CA 94305 USA

[3]Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA 02129 USA

[4]Biological and Biomedical Sciences Program, Harvard Medical School, Boston, MA 02115 USA

[5]Department of Medicine, Stanford University School of Medicine, Stanford CA 94305 USA

[6]Department of Pathology, Harvard Medical School, Boston, MA 02115 USA

*These authors contributed equally to this work.
[#]Corresponding authors

Reprinted with permission by AlphaMed Press from:

Sebastiano V, Maeder ML, Angstman JF, Haddad B, Khayter C, Yeo DT, Goodwin MJ, Hawkins JS, Ramirez CL, Batista LF, Artandi SE, Wernig M, Joung JK. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*. 2011 Nov;29(11):1717-26.

Vittorio Sebastiano derived and validated the iPS cell lines (Figure 2.1). Morgan Maeder engineered, and did the initial testing of, the zinc finger nucleases (Figure 2.2). All gene targeting experiments and characterization of targeted clones (Figures 2.3 and 2.4) were done collaboratively by Morgan Maeder and Vittorio Sebastiano. Morgan Maeder designed and constructed donor templates. Morgan Maeder, James Angstman and Vittorio Sebastiano performed transfections and isolated iPS cell clones. Morgan Maeder and James Angstman characterized clones by PCR, sequencing and Southern blotting. Vittorio Sebastiano performed pluripotency assays on targeted clones.

Abstract

The combination of induced pluripotent stem (**iPS**) cell technology and targeted gene modification by homologous recombination (**HR**) represents a promising new approach to generate genetically corrected, patient-derived cells that could be used for autologous transplantation therapies. This strategy has several potential advantages over conventional gene therapy including eliminating the need for immunosuppression, avoiding the risk of insertional mutagenesis by therapeutic vectors, and maintaining expression of the corrected gene by endogenous control elements rather than a constitutive promoter. However, gene targeting in human pluripotent cells has remained challenging and inefficient. Recently, engineered zinc finger nucleases (**ZFN**s) have been shown to substantially increase HR frequencies in human iPS cells, raising the prospect of employing this technology to correct disease-causing mutations. Here we describe the generation of iPS cell lines from sickle cell anemia patients and *in situ* correction of the disease-causing mutation using three ZFN pairs made by the publicly available Oligomerized Pool Engineering (**OPEN**) method. Gene-corrected cells retained full pluripotency and a normal karyotype following removal of reprogramming factor and drug-resistance genes. By testing various conditions, we also demonstrated that HR events in human iPS cells can occur as far as 82 bps from a ZFN-induced break. Our approach delineates a roadmap for using ZFNs made by an open-source method to achieve efficient, transgene-free correction of monogenic disease mutations in patient-derived iPS cells. Our results provide an important proof of principle that ZFNs can be used to produce gene-corrected human iPS cells that could be used for therapeutic applications.

Introduction

The development of induced pluripotent stem (**iPS**) cell technology has raised prospects for patient-specific therapies of various human diseases.[1, 2] Human iPS cells can grow to virtually unlimited numbers, can differentiate to all somatic tissues, and can be derived from a variety of somatic cells such as fibroblasts obtained by a simple skin biopsy.[3-6] For treatment of monogenic diseases, patient-derived iPS cells might be corrected, differentiated into therapeutically relevant cells, and transplanted back into the patient for restoration of function. A critical component of this envisioned therapeutic strategy is the efficient genetic correction of the disease-causing gene mutation in the iPS cells. One strategy for correction is to use homologous recombination (HR) with an exogenous DNA donor template to modify specific genomic sequences. This *in situ* approach would retain regulation by endogenous genomic elements, a key advantage for genes that must be expressed in a staged fashion through differentiation or development. In a previous report, we used gene targeting in murine iPS cells to correct the sickle cell anemia mutation in a mouse model of this disease.[7] However, in contrast to mouse cells, efficiencies of gene targeting in human iPS and embryonic stem cells are low[8-12] or require the construction of large donor templates,[13, 14] thereby rendering this approach difficult to practice.

Engineered zinc finger nucleases (**ZFN**s) can be used to substantially increase the rate of gene targeting at specific genomic loci.[15-17] ZFNs are customizable restriction enzymes that consist of an engineered zinc finger array fused to a non-specific DNA cleavage domain from the *Fok*I endonuclease.[18] Introduction of a ZFN-induced DNA double-strand break (**DSB**) substantially stimulates gene targeting mediated by homologous recombination (**HR**) with a

"donor template" at the cleavage site. Various publicly available[19-24] and proprietary methods[25] exist to engineer the zinc finger arrays required for targeting the *Fok*I cleavage domain to a specific genomic locus. We recently described a rapid, "open-source" ZFN engineering method known as "Oligomerized Pool Engineering" (**OPEN**)[22] and have used ZFNs made by this approach to modify endogenous genes in zebrafish, plants, and various human cells.[11, 22, 26-28] Previous work in human iPS cells has shown that ZFNs can be used to enhance the efficiency of HR-mediated reporter gene or drug-resistance marker insertion into ZFN-induced DSBs.[9, 11] One recent report also described insertion of a wild-type gene into a ZFN-induced break induced at a potential "safe harbor" locus in human iPS cells derived from patients with X-linked chronic granulomatous disease, thereby complementing the gene defect.[29] However, *in situ* gene correction by ZFNs is a more challenging problem than simple reporter gene insertion because the currently limited targeting range of publicly available ZFN engineering methods can make it difficult to introduce a DSB directly at the site of the mutation. In addition, high efficiency ZFN-mediated HR in human iPS cells requires co-integration of a drug-resistance gene into a nearby intron, thereby necessitating the introduction of transgenic sequences some distance away from the ZFN cleavage site.

Here we demonstrate facile and efficient correction of the sickle cell anemia mutation in patient-derived human iPS cells using ZFNs engineered by the publicly available Oligomerized Pool ENgineering (**OPEN**) method.[22, 23] We accomplished this by using OPEN ZFNs to simultaneously correct the mutation and insert a drug-resistance cassette into a neighboring intron up to 82 base pairs away from the ZFN cut site. Following excision of the reprogramming factor and drug-resistance cassettes, corrected and transgene-free iPS cells retained full

pluripotency and normal karyotypes. Our successful genetic correction of a disease-causing mutation in human iPS cells provides an important proof-of-principle for therapeutic applications and delineates a roadmap for using ZFNs as an efficient tool for targeted manipulation of human iPS cells.

Results

Derivation and characterization of iPS cells from two sickle cell anemia patients

Many different non-integrating methods have been developed for the safe generation of human iPS cells.[5, 33-37] While these methods have great potential, we employed an integrating lentivirus that contains a polycistronic cassette encoding the four required reprogramming factors Oct4, Sox2, Klf4, and c-myc that is flanked by loxP sites.[30, 38] This approach enables high reprogramming efficiencies together with the generation of transgene-free iPS cells following Cre recombinase-mediated excision of the reprogramming cassettes.

Primary fibroblast cultures were established using skin punch biopsies from two patients diagnosed with sickle cell anemia (hereafter Patient 1 and 2). DNA sequencing revealed that Patient 1 has a sickle cell E6V mutation in one beta-globin allele and a splice donor site mutation at the end of exon 1 in the other allele while Patient 2 is homozygous for the E6V mutation (data not shown). Fibroblasts from each patient were infected with a polycistronic reprogramming factor virus and seeded onto mitomycin C-treated feeder cells (Materials and Methods). Multiple colonies appeared after 4 to 6 weeks and these were picked and further expanded in separate dishes. Four lines (hereafter I and II for the clones derived from Patient 1 and III and IV for the clones derived from Patient 2) were selected for continued expansion.

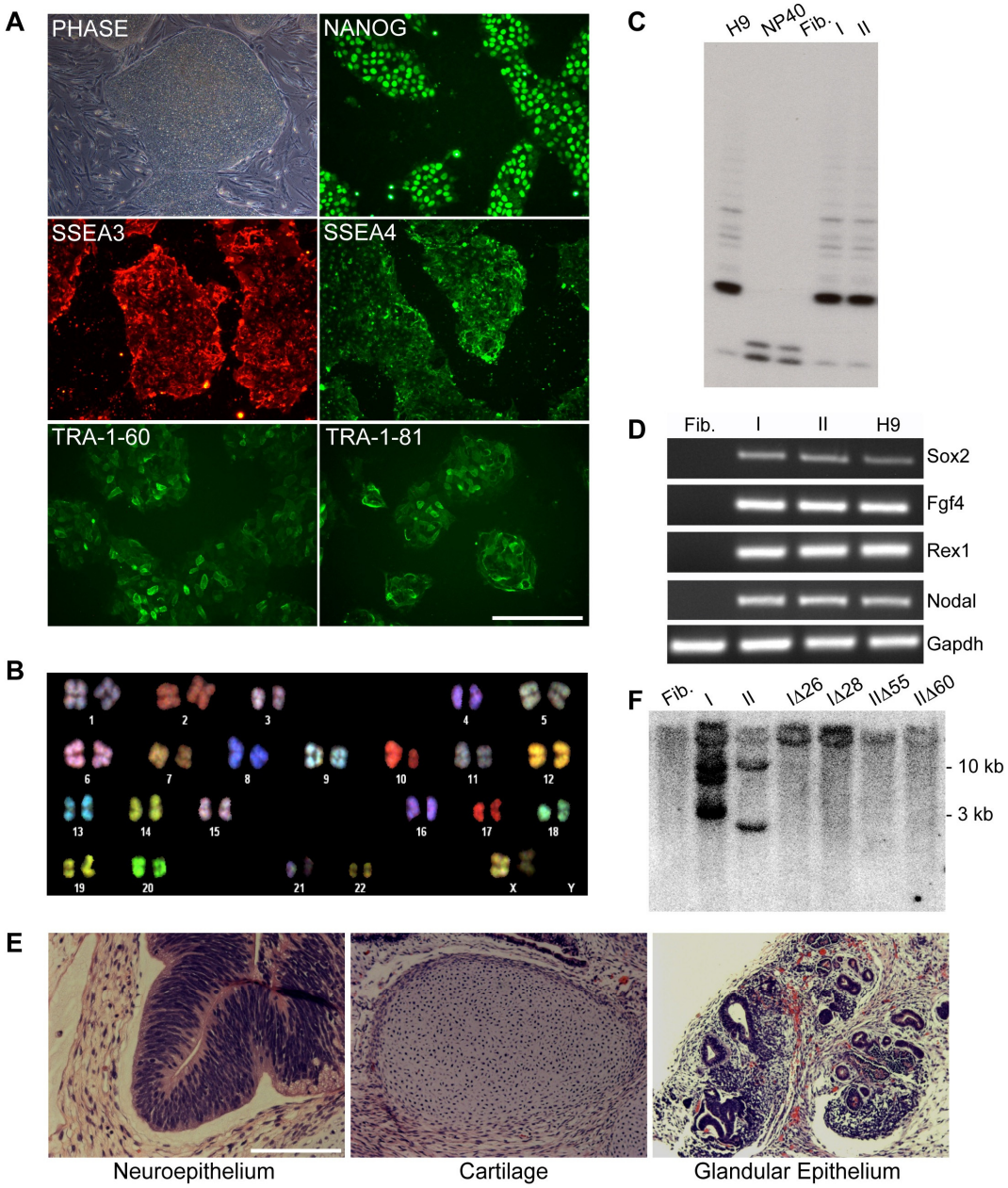


Figure 2.1. Characterization of sickle cell iPS clones I and II. A. Phase contrast image of iPS clone I grown on feeder cells and immunofluorescence staining images of feeder-free cultures. **B.** SKY-FISH analysis of a representative normal diploid metaphase spread of iPS clone I. **C.** Telomerase activity assay: Fibroblasts show no telomerase activity as in the negative control (NP40); conversely iPS clones I and II show a pattern similar to the one observed in H9 ES cells (positive control). **D.** RT-PCR analysis of pluripotency-associated genes in patient parental fibroblasts, iPS clones I and II, and H9 ES cells. **E.** H&E staining of teratoma derived from iPS clone I showing tissues representative of all three germ layers. **F.** Southern Blot analysis using a c-myc probe showing the number of STEMCCA cassette integrations in parental iPS clones I and II and four derivative clones transfected with Cre recombinase. Clones I and II show 3 and 2 integrations, respectively, while clones transfected with Cre recombinase (IΔ26, IΔ28, IΔ55, and IΔ60) show no integrations, similar to what is observed in uninfected fibroblasts.
Abbreviations: Fib.: Fibroblasts from sickle cell anemia patient 1; I: iPS clone I; II: iPS clone II; NP40: negative control for telomerase activity.

After several passages on feeders, these clones could also be successfully propagated under feeder-free conditions (mTeSR1 medium on matrigel-coated plates). All four clones expressed multiple pluripotency markers at the mRNA and protein level, showed telomerase activity similar to control human H9 embryonic stem cells, possessed a normal karyotype, and could be differentiated *in vitro* into derivatives of all three germ layers (Figure 2.1A-D and Supplementary Table 2.1).

Clones I, II and III were injected under the skin of immuno-compromised mice and formed tumors containing various tissue types representing cells from all three germ layers (Figure 2.1E, Supplementary Figure 2.1 and Supplementary Table 2.1). Thus, surprisingly, the lentiviral transgenes appeared to be silenced allowing the cells to differentiate. Southern blot analysis revealed the presence of three proviral integrations in clone I, two in clone II, one in clone III and one in clone IV (Figure 2.1F and Supplementary Table 2.1).

We next verified that the iPS cell lines we generated were stable in the absence of the reprogramming factors. To remove the proviral sequences, clones I and II were transfected with a plasmid encoding Cre recombinase and the puromycin-resistance gene and then selected in the presence of puromycin for 48 hours. Resulting colonies were picked, expanded and their genomic DNA was isolated. A PCR assay detecting specific sequences within the viral vector was performed to screen for potential clones in which all of the reprogramming factors had been removed (data not shown). Two subclones each for clones I and II that were negative for reprogramming factor sequences by PCR were analyzed by Southern blotting to verify loss of the integrated vector sequences (clones IΔ26 IΔ28, IIΔ55, IIΔ60) (Figure 2.1F). All four clones

retained a fully undifferentiated state as judged by the expression of a battery of pluripotency markers and a normal karyotype (Supplementary Figure 2.1 and Supplementary Table 2.2).

Engineering of OPEN ZFNs targeted to the human beta-globin gene

Using our web-based Zinc Finger Targeter (ZiFiT) program,[39] we identified two target sites occurring near the sickle cell anemia mutation in our patient-derived iPS cell clones for which ZFNs could be potentially engineered using the publicly available Oligomerized Pool ENgineering (OPEN) protocol previously described by our group (Figure 2.2A).[23] These two sites fall within 25 bps of the sickle mutation (Figure 2.2B): one is present ~24 bps downstream (ZFN target site 1) and the other is directly at the site of the mutation (ZFN target site 2). OPEN selections for ZFN target sites 1 and 2 were performed and successfully yielded multiple different zinc finger arrays for all four half-sites. Quantitative testing in a bacterial two-hybrid (B2H) reporter assay revealed that the majority of these arrays activated transcription by three-fold or more (data not shown), a threshold we have previously shown is a strong predictor for efficient function as ZFNs in human cells.[11, 22] Based on these results, we chose four and two ZFN pairs for target sites 1 and 2, respectively, to carry forward for additional testing. The sequences of the recognition helices and the activities of these zinc finger arrays in the B2H reporter system are shown in Figure 2.2C.

To test the cleavage activities of our OPEN-selected zinc finger arrays as ZFNs in human cells, we assessed their capabilities to induce site-specific insertion/deletion mutations (**indels**) via non-homologous end joining. Previous work has shown that, in the absence of a donor template, active ZFNs will induce indels with high efficiency at their site of cleavage in somatic

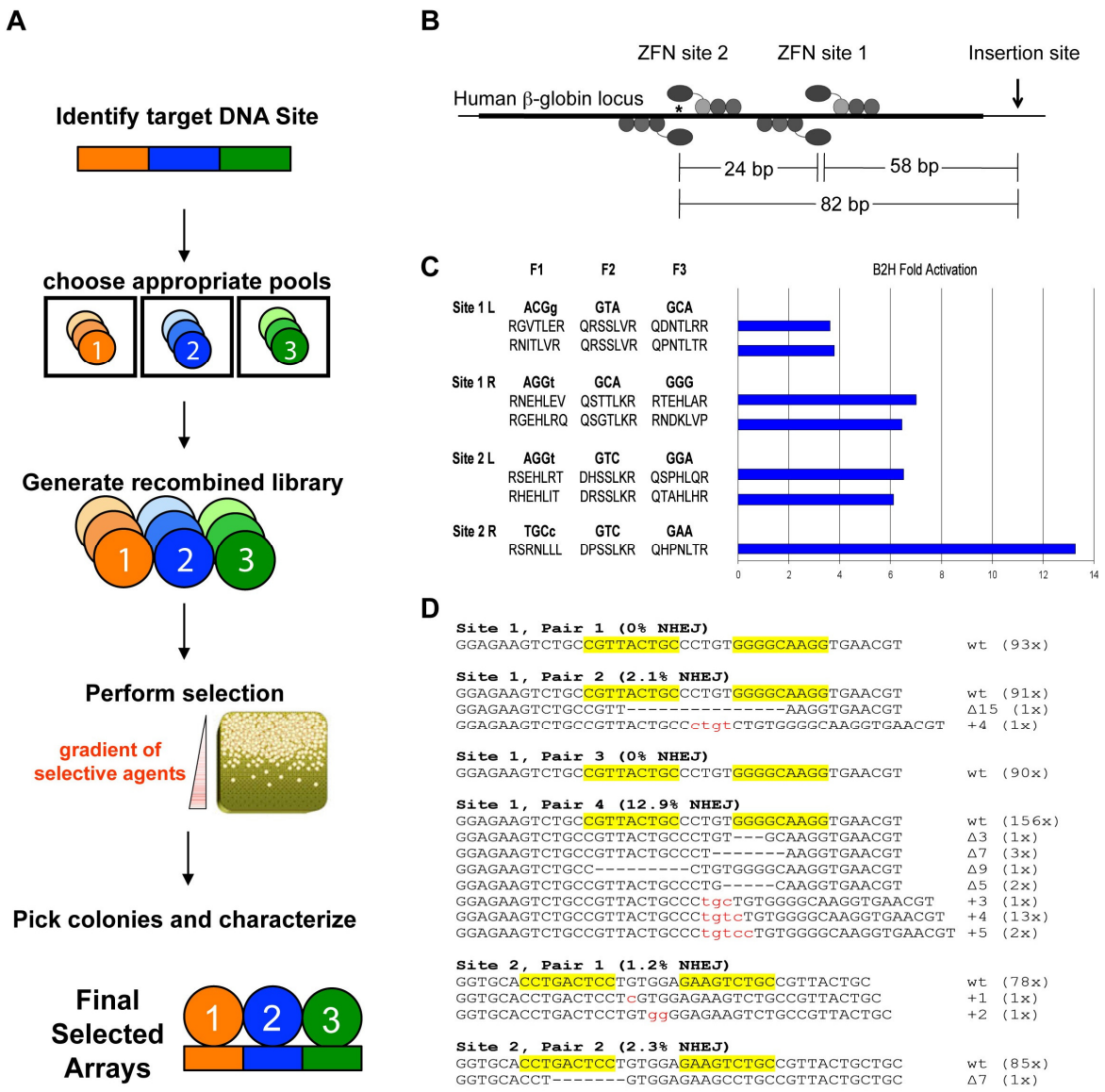


Figure 2.2. Engineering and characterization of ZFNs

A. Schematic overview of Oligomerized Pool Engineering (OPEN) strategy for constructing zinc finger arrays. Individual zinc fingers are depicted as spheres and DNA target sites as rectangles. **B.** Schematic of the human β-globin locus showing the position of the two ZFN target sites in relation to the sickle cell anemia mutation (asterisk) and to the insertion site of the selectable marker (arrow). Exon represented by thick line, introns by thin lines. Not drawn to scale. **C.** Recognition helix sequences and bacterial two-hybrid (B2H) fold-activation values of selected zinc finger proteins engineered by OPEN. Amino acid sequence of the recognition helices are shown below their respective three base pair target DNA sites. B2H fold-activation values are shown as blue bars on the right. **D.** Cleavage efficiency of ZFN pairs as measured by induction of mutations caused by non-homologous end-joining (NHEJ) repair in HEK293 cells (for site 1) or primary human fibroblasts (for site 2). Wild-type sequence is shown with zinc finger binding sites highlighted in yellow. Mutant sequences are shown below with deletions indicated by dashes and insertions shown in red, lower-case letters. The frequency with which each sequence was detected is indicated to the right.

human cells.[40] For ZFN target site 1, two of the four pairs tested showed activity in human HEK293 cells, with one pair (hereafter designated ZFN Pair A) exhibiting a substantially higher mutation frequency (12.9%) than the others (Figure 2.2D). For ZFN target site 2, both pairs (hereafter ZFN Pairs B and C) tested yielded roughly equivalent mutation frequencies in primary human fibroblasts derived from Patient 1 (Figure 2.2D). Based on these experiments, we carried forward ZFN pairs A, B, and C for gene correction experiments in our sickle cell anemia iPS cells.

Construction of a "donor template" required for ZFN-mediated correction of the sickle cell anemia mutation

Our ZFN-based strategy to correct the sickle cell anemia mutation also requires a homologous "donor template" for introducing the desired correction. We therefore constructed a donor template that harbors ~1.6 kb of beta-globin gene sequence centered approximately on the location of the sickle cell anemia E6V mutation but which codes for the wild-type E6 codon (Figure 2.3A). In addition, we introduced two translationally silent mutations into the donor construct to reduce binding and cleavage of the donor by ZFN Pair A. Based on previous experience,[9, 11] we presumed that ZFN-induced gene targeting in iPS cells would require drug-selection to identify desired recombinants. Therefore, we modified the donor template by inserting a puromycin or neomycin-resistance gene cassette into the intron located downstream of the E6V mutation. The drug-resistance cassette was flanked by loxP sites to enable its subsequent excision using Cre recombinase. In this configuration, the 5' end

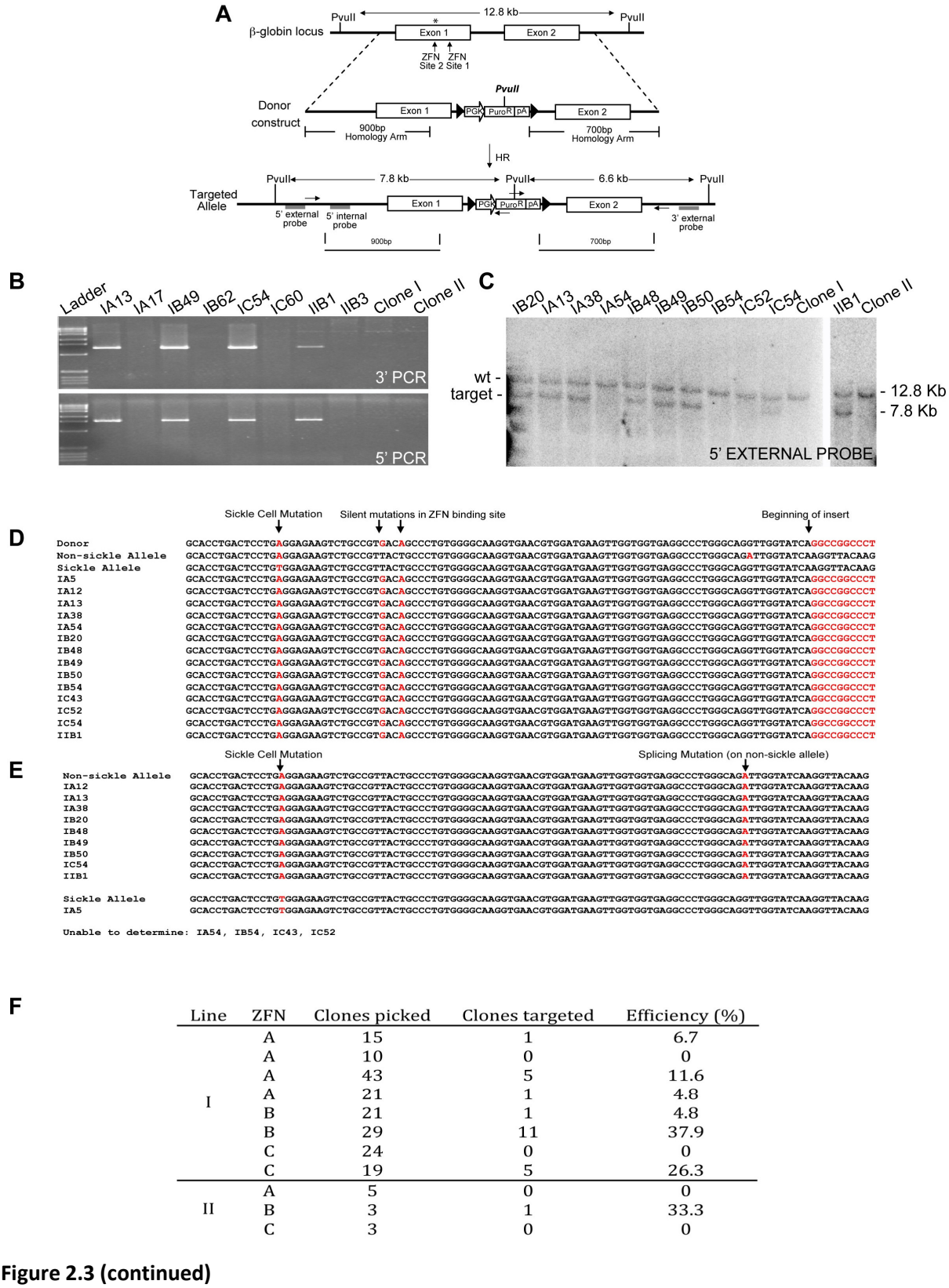
of the puromycin or neomycin-resistance gene insertion is positioned ~58 or ~82 bps downstream of the cleavage sites for ZFN Pairs A or B/C, respectively (Figures 2.2B and 2.3A).

Efficient correction of the sickle cell anemia mutation in human iPS cells using ZFN-mediated gene targeting

To perform gene correction in our sickle cell iPS cells, we transfected plasmids expressing each of the ZFN pairs together with the donor template into clone I (Materials and Methods). We performed two to four independent experiments for each of the three ZFN Pairs (A, B, and C) and obtained puromycin-resistant colonies for all transfections (Figure 2.3F). We manually picked and expanded drug-resistant clones, isolated genomic DNA, and performed PCR reactions to screen for the presence of the puromycin-resistance gene in the endogenous beta-globin gene using independent sets of primers expected to amplify the 5' and the 3' junctions in a correctly targeted beta-globin locus (Figure 2.3B). We achieved targeting efficiencies as high as 37.9% (mean of 9.8 %) as judged by this PCR screening assay (Figure 2.3B and 2.3F). DNA sequencing of 13 PCR-positive lines revealed the presence of a successfully targeted allele in all of these cells (Figure 2.3D). The presence of a polymorphism on the non-sickle allele allowed us to identify which of the two alleles had undergone targeting in each clone: seven of the 13 candidates (IA5, IA12, IA13, IA38, IB20, IB49 and IC54) were clonal populations with all but one (IA5) having undergone gene targeting of the sickle cell allele while six of the 13 candidates were mixed populations containing both targeted and non-targeted cells (IA54, IB48, IB50, IB54, IC43 and IC52). Following further expansion in culture, these six

Figure 2.3. Gene targeting of the endogenous human beta-globin locus

A. Schematic overview of gene targeting strategy for the human beta-globin locus. The sickle cell mutation is labeled by an asterisk and ZFN target sites are indicated by arrows. The desired recombination event inserts a PGK promoter-puromycin-resistance cassette or PGK promoter-neomycin-resistance cassette flanked by loxP sites (black triangles) into the intron between exons 1 and 2. Southern blot probes are indicated by grey bars and PCR primers are indicated by arrows. Not drawn to scale. **B.** Representative PCR analysis of puromycin-resistant clones. PCR of gene targeted clones gives a band of 1.4 kb when amplified with 3' primers and 1.6 kb with 5' primers. Non-targeted clones fail to give PCR product. **C.** Southern blot using a 5' external probe on genomic DNA digested with *Pvu*II. A beta-globin allele that has not undergone gene targeting gives a 12.8kb band while a targeted allele gives a 7.8kb band due to the presence of a *Pvu*II site in the puromycin-resistance gene. Note that after expansion in culture, four clones previously revealed to be mixed populations by sequencing no longer possess the gene targeted allele. **D.** Sequence of gene targeted allele detected by PCR assay. The allele that has undergone homologous recombination with the donor template is wild-type at the E6 codon that is mutated in sickle cell disease, contains translationally silent mutations incorporated into ZFN target site 1, and has incorporated the puromycin or neomycin-resistance cassette. **E.** Sequence of the non-gene targeted allele. The untargeted allele in clones IA12, IA13, IA38, IB20, IB48, IB49, IB50, IC54 and IIBI is the allele that does not harbor the sickle cell mutation, indicating that the sickle cell allele underwent gene targeting. In clone IA5, the untargeted allele is the sickle cell allele. It was not possible to determine which allele underwent gene targeting in clones IA54, IB54, IC43 or IC52 presumably because these were mixed populations of both targeted and untargeted clones. **F.** Table showing the number of puromycin or neomycin-resistant colonies picked for each experiment and the percentage of these that were determined to be correctly targeted by the PCR screening assay. Asterisk indicates that ZFNs were expressed as fusions to the heterodimeric ELD/KKR FokI nuclease domain as opposed to the heterodimeric EL/KK FokI nuclease domain (no asterisk).



F

Line	ZFN	Clones picked	Clones targeted	Efficiency (%)
I	A	15	1	6.7
	A	10	0	0
	A	43	5	11.6
	A	21	1	4.8
	B	21	1	4.8
	B	29	11	37.9
	C	24	0	0
	C	19	5	26.3
II	A	5	0	0
	B	3	1	33.3
	C	3	0	0

Figure 2.3 (continued)

mixed populations resolved into clonal lines: two consisting of cells successfully targeted in the sickle E6V allele (IB48 and IB50) as judged by DNA sequencing (Figure 2.3D and 2.3E) and four consisting of non-targeted cells (IA54, IB54, IC43 and IC52) as determined by PCR analysis (Supplementary Table 2.3). Importantly, it is worth noting that DNA sequence analysis showed that for all of the corrected clonal lines the non-targeted allele did not show evidence of ZFN-mediated indels introduced by non-homologous end joining (Figure 2.3E).

To provide further evidence of successful homologous recombination at the beta-globin locus, we also performed Southern blots on seven correctly targeted clones (IA13, IA38, IB20, IB48, IB49, IB50, IC54) using both 5' external and 3' external probes. These experiments confirmed stable presence of the puromycin-resistance insert in the beta-globin locus (Figure 2.3C and data not shown). In addition, Southern blots performed with these same probes on three of the four clones that had become negative for the gene targeting event as judged by PCR confirmed that these cells (IA54, IB54, and IC52) no longer contain a targeted clone following their expansion in culture (Figure 2.3C). Finally, Southern blots performed using a 5' homology arm (internal) probe confirmed the absence of random insertions of the donor template in all seven correctly targeted clones (Supplementary Figure 2.2B).

We next sought to demonstrate whether ZFN-induced gene correction could be performed in additional iPS cell lines. We therefore isolated puromycin or neomycin-resistant colonies from targeting experiments with iPS cell lines II and IV and identified successful gene correction in both lines, as judged by PCR, Southern blot and DNA sequencing (clone IIB1 in Figure 2.3B-F; clones IVA3 and IVC1 in Figure 2.3F and Supplementary Figure 2.2A). DNA

sequencing of the non-targeted allele in all three of these clones also revealed the absence of ZFN-induced indel mutations (Figure 2.3E and data not shown). Taken together, our results with iPS cell clones I, II and IV demonstrate that all three ZFN Pairs A, B, and C can efficiently induce stable gene correction of the sickle cell mutation in human iPS cells without inducing additional mutagenesis of the other beta-globin allele in the same cell.

Evaluation of ZFN off-target effects at other closely related globin genes

An important question to address for eventual therapeutic application of ZFNs is the range of unwanted "off-target" mutagenic effects caused by these reagents [16]. For treatment of sickle cell disease, it is particularly important to ensure that the highly related paralogous genes encoding gamma-globin and delta-globin are not inadvertently mutagenized by the ZFNs targeted to beta-globin. The single delta-globin and two gamma-globin genes each possess DNA sequences that are very similar to the target sites for ZFN pairs A, B, and C (Supplementary Figure 2.3A). DNA sequencing revealed that none of these sites were altered in any of the nine corrected iPS cell clones derived from the three different ZFN pairs in iPS lines I and II (Supplementary Figure 2.3A). Additionally, we computationally identified 12 sites in human genome sequence that differed from the ZFN target sites by only a single base pair change (Supplementary Methods). We reasoned that these 3 sites for ZFN A and 9 sites for ZFNs B and C would be the most likely off-target sites for ZFN-induced mutations. Sequencing of these off-target sites in three targeted clones (IA13, IA38 and IB49) showed that none of them harbored ZFN-induced mutations (Supplementary Figure 2.3B). These results show that our ZFNs did not affect very closely related off-target sites in both paralogous genes and the most closely related

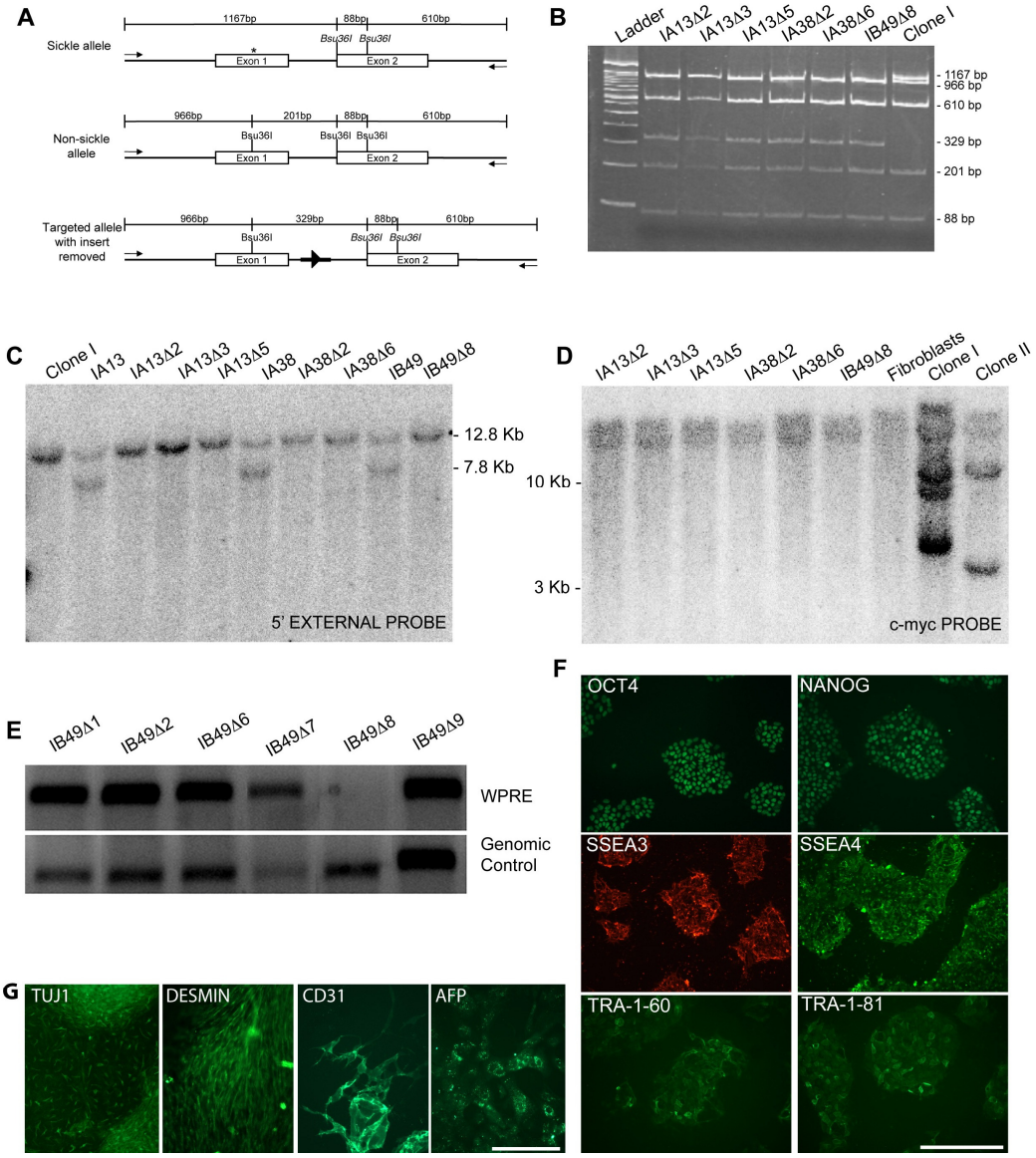


Figure 2.4. Characterization of gene targeted iPS cell clones

A. Schematic illustrating location of primers and restriction sites in the beta-globin locus used to identify gene-targeted clones that have undergone removal of the puromycin-resistance cassette. PCR primers external to donor template homology arms are depicted by arrows. The restriction enzyme *Bsu36*I will cleave sequence with a wild-type E6 codon but not sequence containing the E6V sickle cell mutation (asterisk). Residual "scar" sequence containing a single loxP site (black triangle) and flanking restriction sites (bold dark line) are left in the intron sequence. Expected sizes of bands are labeled. Not drawn to scale. **B.** PCR and restriction digest analysis for removal of puromycin-resistance insert. PCR products were digested with *Bsu36*I and run on a 5% polyacrylamide gel to yield fragments of expected sizes as depicted in A. **C.** Southern blot analysis to confirm removal of the puromycin-resistance insert. Blot was performed using an external probe that hybridizes upstream of the 5' homology arm on genomic DNA digested with *Pvu*II (as show in Figure 2.3A). **D.** Southern blot analysis to confirm removal of reprogramming factors from gene targeted clones. Blot was performed using a probe that hybridizes to the c-myc gene on genomic DNA digested with *BamH*I. Genomic DNA from the sickle cell anemia patient primary fibroblasts was used as a control to define the endogenous c-myc allele. Note that the last three lanes are the same as the first three lanes in Figure 2.1F. **E.** PCR screening assay for the WPRE sequence to verify successful excision of the STEMCCA reprogramming cassette **F.** Immunofluorescence images of feeder-free cultures (clone IA13Δ2) stained for pluripotency-associated transcription factors and surface markers. **G.** Immunofluorescence images of clone IA13Δ2 differentiated *in vitro*.

potential off-target sites in the genome during the process of correction, demonstrating the high degree of specificity these nucleases possess for their intended target sites.

Generation of transgene-free, gene-corrected iPS cell lines

We next sought to simultaneously excise both the floxed puromycin-resistance gene incorporated into the beta-globin intron and the randomly integrated floxed reprogramming factors by transiently expressing Cre recombinase in our corrected iPS cells. Three independent corrected lines derived from Clone I (IA13, IA38, and IB49) were infected with an adenovirus expressing a Cre recombinase-GFP fusion protein. Single cells with high level GFP expression were isolated by FACS 48 hours post-infection and plated. After two to three weeks, single colonies were picked and clonally expanded. PCR and restriction digest with the enzyme *Bsu36*I, which cleaves the wild-type but not sickle cell sequence, as well as DNA sequencing, confirmed excision of the puromycin-resistance cassette and the presence of the expected residual 34 bp loxP site "scar" with several associated restriction site sequences in the intron of the corrected beta-globin gene for six clones (IA13Δ2, IA13Δ3, IA13Δ5, IA38Δ2, IA38Δ6, and IB49Δ8) derived from the three corrected lines (Figures 2.4A, 2.4B and data not shown). Southern blotting verified excision of the puromycin-resistance cassette from these six clones (Figure 2.4C). PCR and Southern blot also demonstrated removal of the reprogramming factor cassette for these six clones (Figures 2.4D, 2.4E and data not shown). These successfully corrected, transgene-free iPS cell lines remained pluripotent as judged by expression of multiple pluripotency markers and *in vitro* differentiation into derivatives of the three germ layers (Figure 2.4F, 2.4G and Supplementary Figure 2.4). Finally, SKY-FISH analysis showed that

all six correctly targeted clones were karyotypically normal (Supplementary Table 2.2). Thus, no translocations were observed despite transient ZFN and Cre recombinase expression in our cell lines harboring multiple loxP sequences.

Discussion

In this study, we used publicly available engineered ZFN technology to effect facile, efficient, and stable correction of the E6V mutation in the beta-globin gene in three independent iPS cell lines derived from two patients with sickle cell anemia. Our data show that ZFN technology allows efficient gene targeting of disease-causing mutations with no evidence for off-target cleavage in closely related DNA sequences. Our results demonstrate that reprogramming and ZFN-induced gene targeting can be used in combination to create corrected patient-derived iPS cells and provide an important initial proof-of-principle that such an approach might be used in the future to develop therapies for monogenic diseases.

In contrast to previous studies including those with human iPS cells,[9, 11, 29, 41] we introduced transgenic sequences up to 82 base pairs away from the ZFN cut site – a prerequisite for correcting disease-causing mutations. Indeed, previous work in mouse ES cells has demonstrated that the efficiency of alterations introduced by homologous recombination decreases substantially as distance from the nuclease-induced DSB increases.[42, 43] Our studies illustrate that several configurations of the ZFN-induced DSB, the genetic correction event, and the drug-resistance cassette insertion point can be successfully used when performing ZFN-enhanced HR in human iPS cells. In one case, the ZFN-induced DSB (for ZFN target site 2) directly overlapped the mutation to be corrected with the drug-resistance cassette insertion

positioned 82 bps downstream. This demonstrates that ZFNs can mediate cassette insertion even when it occurs nearly 100 bps away from the DSB. In a second case, the ZFN-induced DSB was positioned between the correction event and insertion point -- 24 bps downstream of the correction and 58 bps upstream of the insertion. Success in this latter configuration shows that two events, a correction and an insertion, can occur efficiently even when they are positioned some distance away from each other and on opposite sides of the DSB. These results allow us to draw the important conclusion that the gene correction event and the drug-resistance cassette insertion point do not need to be positioned precisely at the location of the ZFN-induced DSB.

A particularly important area for future investigation will be to define the full extent of genetic alterations introduced into iPS cells during generation and repair. Karyotype analyses showed that the reprogramming, ZFN-induced gene correction and Cre recombinase excision processes we performed did not introduce any gross chromosomal translocations or alterations into the patient-derived iPS cells. However, we know that one or more loxP "scar" sequences were left behind at undefined locations in the genome following excision of reprogramming factors by Cre recombinase. Furthermore, following correction of the sickle cell mutation, we also recognize that one residual loxP scar remains in the first intron of the beta-globin gene. Presumably, additional somatic mutations were either present in fibroblasts of the biopsied skin or introduced during the reprogramming process as suggested by recent studies examining genome-wide coding sequences and copy number variations [44, 45]. Finally, although it remains theoretically possible that additional "off-target" mutations may have been introduced by the ZFNs, the absence of ZFN-induced indel mutations at the uncorrected beta-globin allele (with a

perfect ZFN target site) as well as at highly similar sites in the paralogous delta- and gamma-globin genes and at closely matched, computationally identified potential off-target sites in any of our gene-corrected iPS clones strongly suggests that off-target effects are likely to be minimal.

Previous studies have used other methods such as adeno-associated virus (AAV) vector-based,[46, 47] bacterial artificial chromosome-based,[13, 48] and helper-dependent adenovirus-based[14] donor templates to enhance gene-targeting efficiencies in human iPS cells. Although these technologies, like ZFNs, improve the targeting efficiency, the absolute efficiencies of successful recombination events are still sufficiently low enough to require the use of drug selection. In addition, the construction of the donor templates for these methods is more complicated than those required for ZFN-induced HR. However, we note that any of these three methods could also, in principle, be used together with ZFNs to effect correction in iPS cells. It will be of interest to determine if combined approaches could lead to further enhancements in the efficiencies of gene correction in human iPS cells. We also note that a recent report demonstrated the successful use of ZFNs together with donors lacking drug-selection cassettes (both plasmids and single-stranded oligonucleotides) to effect efficient gene correction or gene editing in human pluripotent stem cells.[49] However, the efficiencies of gene alteration in these experiments were notably lower than those performed using donors harboring a drug-selection cassette.

Our efficient and reproducible correction of the sickle cell mutation in human iPS cells provides an important proof-of-principle and defines a framework for the use of engineered ZFN technology to correct a disease-causing mutation by HR. This result advances the field one

important step closer to clinical realization of combined gene and cell therapy based on human iPS cells especially if such methods are used in combination with integration-free methods (e.g. mRNA- and or protein-based approaches) for the derivation of iPSCs.[35, 37]

Materials and Methods

Detailed Methods are provided in the Supplementary Materials.

Derivation and Culture of iPS cells

Derivation, modification, and analysis of patient-derived iPS cells were performed under IRB protocols approved by Stanford University and Massachusetts General Hospital. Dermal fibroblasts from two sickle cell anemia patients were infected with the polycistronic STEMCCA lentiviral reprogramming vector. [30] 10^5 fibroblasts were seeded in MEF medium and infected 24 hours later. After 6 days, cells were transferred onto inactivated MEFs. The following day MEF medium was replaced with hES medium and the cells were grown for up to 8 weeks until hES-like colonies started to emerge. iPS colonies were manually picked and expanded on MEFs. After approximately 10 passages the clones were transferred to feeder-free culture conditions in mTeSR1 medium (Stem Cell Technologies).

Engineering ZFNs targeted to the human β-globin gene

Zinc finger arrays binding each of these target sites were selected using either the previously described OPEN method[22] or a modified protocol that uses antibiotic selection.[24] The OPEN selections we performed required the use of several additional OPEN pools for new 3

bp subsites (J.K. Joung et al., manuscript in preparation). Candidates from OPEN selections were assayed for DNA-binding activity in the bacterial-two-hybrid system as previously described.[23] Selected zinc fingers were cloned on an *Xba*I-*BamH*I fragment into a ZFN expression vector that expresses heterodimeric ZFNs (either EL/KK[31] or ELD/KKR[32] fused by a T2A self-cleaving peptide. The CMV promoter driving ZFN expression in the resulting plasmids was then replaced with an EF1α promoter. This ZFN expression plasmid also contains a PGK promoter-GFP reporter gene cassette that can be used to assess transfection efficiency.

Transient transfection for Gene Targeting and Loop-out in iPS cells

For gene targeting, $1x10^6$ iPSCs were transfected with 2 μg of ZFN-encoding plasmid and 10 μg of donor template plasmid by Amaxa nuclefection. After four days, cells were selected in media containing puromycin (1 μg/ml) for 2 weeks. To excise selection and/or reprogramming cassettes, cells were transfected with an EF1a-Cre-ires-puro plasmid (a gift kindly provided by Gustavo Mostoslavsky). Alternatively infection with adenovirus expressing a Cre-recombinase-GFP fusion (Vector Biolabs) was used. FACS was used to isolate GFP-positive cells that were seeded as single cells on Matrigel in mTeSR1 medium containing 2 μM Thiazovivin. Individual colonies were manually picked and screened by PCR.

Southern Blotting

Genomic DNA was separated on a 0.8% agarose gel following an overnight restriction digest with indicated restriction enzymes, transferred to a nylon membrane (Amersham), and

hybridized with ^{32}P-labeled probes as indicated made by random priming following the manufacturer's instructions (Agilent Prime-It II kit).

Teratoma Formation and Analysis

Cells from a confluent 10 cm plate were harvested by Accutase digestion, resuspended in phosphate buffered saline, and injected subcutaneously into immuno-compromised NOG-SCID mice (obtained from the Jackson Laboratories). Four to eight weeks after injection, teratomas were dissected, fixed in 4% PFA, and processed for H&E.

Expression analysis

Immunofluorescence and RT-PCR analysis was performed following standard protocols as described in the Supplementary Materials. Please see Supplementary Methods also for a list of primer sequences and antibodies used in this study.

Karyotyping

Metaphase spreads were prepared after following treatment with 10 ug/mL of Colcemid (Invitrogen) and processed for SKY-FISH hybridization and image analysis according to the manufacturer's instructions (Applied Spectral Imaging).

Acknowledgments

We thank the patients who volunteered to donate a skin biopsy for this project, Rob Cho for consenting the patients and collecting the biopsies, Pei Wang for help with the Adeno-Cre

infections and FACS sorting, Stacey Thibodeau-Beganny for help with plasmid construction, and Gerlinde Wernig for providing critical technical advice. This work was supported by National Institutes of Health R01 GM069906 (J.K.J.) and Pioneer Award DP1 OD006862 (J.K.J.); by National Science Foundation Graduate Research Fellowships (M.L.M. & C.L.R.); by Start-up funds of the Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine (M.W.); by the Donald E. and Delia B. Baxter Foundation (M.W.); and by the CIRM Training Grant for postdoctoral fellows (V.S.).

References

1. Yamanaka S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. Cell Stem Cell. 2007;1:39-49.

2. Saha K, Jaenisch R. Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell. 2009;5:584-595.

3. Lowry WE, Richter L, Yachechko R, et al. Generation of human induced pluripotent stem cells from dermal fibroblasts. Proc Natl Acad Sci U S A. 2008;105:2883-2888.

4. Park IH, Zhao R, West JA, et al. Reprogramming of human somatic cells to pluripotency with defined factors. Nature. 2008;451:141-146.

5. Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861-872.

6. Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007;318:1917-1920.

7. Hanna J, Wernig M, Markoulaki S, et al. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science. 2007;318:1920-1923.

8. Hockemeyer D, Jaenisch R. Gene Targeting in Human Pluripotent Cells. Cold Spring Harb Symp Quant Biol. 2011.

9. Hockemeyer D, Soldner F, Beard C, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. 2009;27:851-857.

10. Irion S, Luche H, Gadue P, et al. Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat Biotechnol. 2007;25:1477-1482.

11. Zou J, Maeder ML, Mali P, et al. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. Cell Stem Cell. 2009;5:97-110.

12. Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells. Nat Biotechnol. 2003;21:319-321.

13. Howden SE, Gore A, Li Z, et al. Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. Proc Natl Acad Sci U S A. 2011.

14. Liu GH, Suzuki K, Qu J, et al. Targeted Gene Correction of Laminopathy-Associated LMNA Mutations in Patient-Specific iPSCs. Cell Stem Cell. 2011.

15. Urnov FD, Rebar EJ, Holmes MC, et al. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 2010;11:636-646.

16. Handel EM, Cathomen T. Zinc-finger nuclease based genome surgery: it's all about specificity. Curr Gene Ther. 2011;11:28-37.

17. Carroll D. Zinc-finger nucleases: a panoramic view. Curr Gene Ther. 2011;11:2-10.

18. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. 1996;93:1156-1160.

19. Bae KH, Kwon YD, Shin HC, et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. Nat Biotechnol. 2003;21:275-280.

20. Mandell JG, Barbas CF, 3rd. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res. 2006;34:W516-523.

21. Carroll D, Morton JJ, Beumer KJ, et al. Design, construction and in vitro testing of zinc finger nucleases. Nat Protoc. 2006;1:1329-1341.

22. Maeder ML, Thibodeau-Beganny S, Osiak A, et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell. 2008;31:294-301.

23. Maeder ML, Thibodeau-Beganny S, Sander JD, et al. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. Nat Protoc. 2009;4:1471-1501.

24. Sander JD, Dahlborg EJ, Goodwin MJ, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods. 2011;8:67-69.

25. Doyon Y, McCammon JM, Miller JC, et al. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol. 2008;26:702-708.

26. Zhang F, Maeder ML, Unger-Wallace E, et al. High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci U S A. 2010;107:12028-12033.

27. Townsend JA, Wright DA, Winfrey RJ, et al. High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature. 2009;459:442-445.

28. Foley JE, Yeh JR, Maeder ML, et al. Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). PLoS ONE. 2009;4:e4348.

29. Zou J, Sweeney CL, Chou BK, et al. Oxidase deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease mediated safe harbor targeting. Blood. 2011.

30. Sommer CA, Stadtfeld M, Murphy GJ, et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells. 2009;27:543-549.

31. Miller JC, Holmes MC, Wang J, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. 2007;25:778-785.

32. Doyon Y, Vo TD, Mendel MC, et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. 2011;8:74-79.

33. Okita K, Nakagawa M, Hyenjong H, et al. Generation of mouse induced pluripotent stem cells without viral vectors. Science. 2008;322:949-953.

34. Yu J, Hu K, Smuga-Otto K, et al. Human induced pluripotent stem cells free of vector and transgene sequences. Science. 2009;324:797-801.

35. Kim D, Kim CH, Moon JI, et al. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell. 2009;4:472-476.

36. Fusaki N, Ban H, Nishiyama A, et al. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci. 2009;85:348-362.

37. Warren L, Manos PD, Ahfeldt T, et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell. 2010;7:618-630.

38. Somers A, Jean JC, Sommer CA, et al. Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. Stem Cells. 2010;28:1728-1740.

39. Sander JD, Maeder ML, Reyon D, et al. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. 2010;38:W462-468.

40. Santiago Y, Chan E, Liu PQ, et al. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A. 2008;105:5809-5814.

41. DeKelver RC, Choi VM, Moehle EA, et al. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res. 2010;20:1133-1142.

42. Donoho G, Jasin M, Berg P. Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol Cell Biol. 1998;18:4070-4078.

43. Stark JM, Pierce AJ, Oh J, et al. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. Mol Cell Biol. 2004;24:9305-9316.

44. Hussein SM, Batada NN, Vuoristo S, et al. Copy number variation and selection during reprogramming to pluripotency. Nature. 2011;471:58-62.

45. Gore A, Li Z, Fung HL, et al. Somatic coding mutations in human induced pluripotent stem cells. Nature. 2011;471:63-67.

46. Mitsui K, Suzuki K, Aizawa E, et al. Gene targeting in human pluripotent stem cells with adeno-associated virus vectors. Biochem Biophys Res Commun. 2009;388:711-717.

47. Khan IF, Hirata RK, Wang PR, et al. Engineering of human pluripotent stem cells by AAV-mediated gene targeting. Mol Ther. 2010;18:1192-1199.

48. Song H, Chung SK, Xu Y. Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. Cell Stem Cell. 2010;6:80-89.

49. Soldner F, Laganiere J, Cheng AW, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell. 2011;146:318-331.

Chapter 3

Robust and Synergistic Regulation of Human Gene Expression Using TALE-Activators

Morgan L. Maeder[1,2], Samantha J. Linder[1,*], Deepak Reyon[1,3,*], James F. Angstman[1], Yanfang Fu[1,3], Jeffry D. Sander[1,3], J. Keith Joung[1,2,3]

[1]Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129

[2]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02115

[3]Department of Pathology, Harvard Medical School, Boston, MA 02115

Reprinted with permission by Nature Publishing Group from:

Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, Joung JK. Robust, synergistic regulation of human gene expression using TALE activators. *Nature Methods*. 2013 Mar;10(3):243-5.

Morgan Maeder designed and oversaw all experiments, designed and constructed TALE-activator expression vectors, selected TALE sites and performed all transfections, RNA isolations and qRT-PCR assays for *NTF3* and microRNA experiments. Deepak Reyon and Yanfang Fu engineered TALE arrays and Samantha Linder performed transfections and ELISA assays for *VEGF-A* experiments.

Abstract

Artificial TAL effector-based transcriptional activators (TALE-activators) offer broad utility but previous studies suggest that these monomeric proteins often possess low activities. Here we demonstrate that TALE-activators can robustly function individually or in synergistic combinations to increase expression of human protein-coding or microRNA genes over wide dynamic ranges. Our findings will encourage use of TALE-activators for research and therapeutic applications and guide design of other novel monomeric TALE-based fusion proteins.

Introduction

Rapid advances in *Xanthomonas*-derived transcription activator-like (TAL) effector technology have enabled any researcher to construct customizable DNA-binding domains with broad potential uses for targeted alteration of gene sequence or expression. Highly conserved 33-35 amino acid TAL effector repeat domains each bind to one nucleotide of DNA with specificity dictated by the identities of two hypervariable residues.[2] This one-to-one code enables the design of proteins with desired DNA-binding specificities by simply joining TAL effector repeats into an array. Recently, considerable effort has been focused on methods for engineering dimeric TAL effector nucleases (TALENs), artificial proteins composed of customized TAL effectors fused to a nuclease domain, that enable routine targeted modification of endogenous genes in a variety of different organisms and cell types.[3] Engineered TAL effectors have also been fused to heterologous transcriptional activation domains to construct artificial monomeric TAL effector activators (TALE-activators)

(Supplementary Figure 3.1).[4-12] These customizable transcription factors have broad potential applications in basic biological research, synthetic biology, and disease therapy.

In contrast to the high efficiencies of dimeric TALENs, the activities of monomeric TALE-activators reported thus far in the literature have often been disappointingly modest at best in their abilities to increase target endogenous gene expression. For example, 22 of 26 published TALE-activators (for which quantitative information is available)[3-6, 8, 9] failed to induce target endogenous gene expression by five-fold or more (Supplementary Table 3.1). Although these results suggest that TALE-activators may not function robustly on endogenous genes, examination of the 26 published TALE-activators reveals that they have great variability in their architectures (e.g. – in the number of TALE repeats, length and sequences of the TALE-derived domains that flank the TALE repeat arrays, transcription activation domain used; Supplementary Table 3.1). The use of multiple different architectures makes it difficult to ascertain which parameters may or may not influence the activities of the various TALE-activators tested in previous studies.

Results

Here we used our recently developed Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) assembly method[13] to systematically test the activities of TALE-activators built on this platform. In initial experiments, we constructed a large series of TALE-activators composed of a variable number of TAL effector repeats and tested their abilities to stimulate expression of the endogenous human *VEGF-A* gene. We targeted nine regions that all lie within a DNase I hyper-sensitive site (HSS) located ~500 bp downstream of the *VEGF-A*

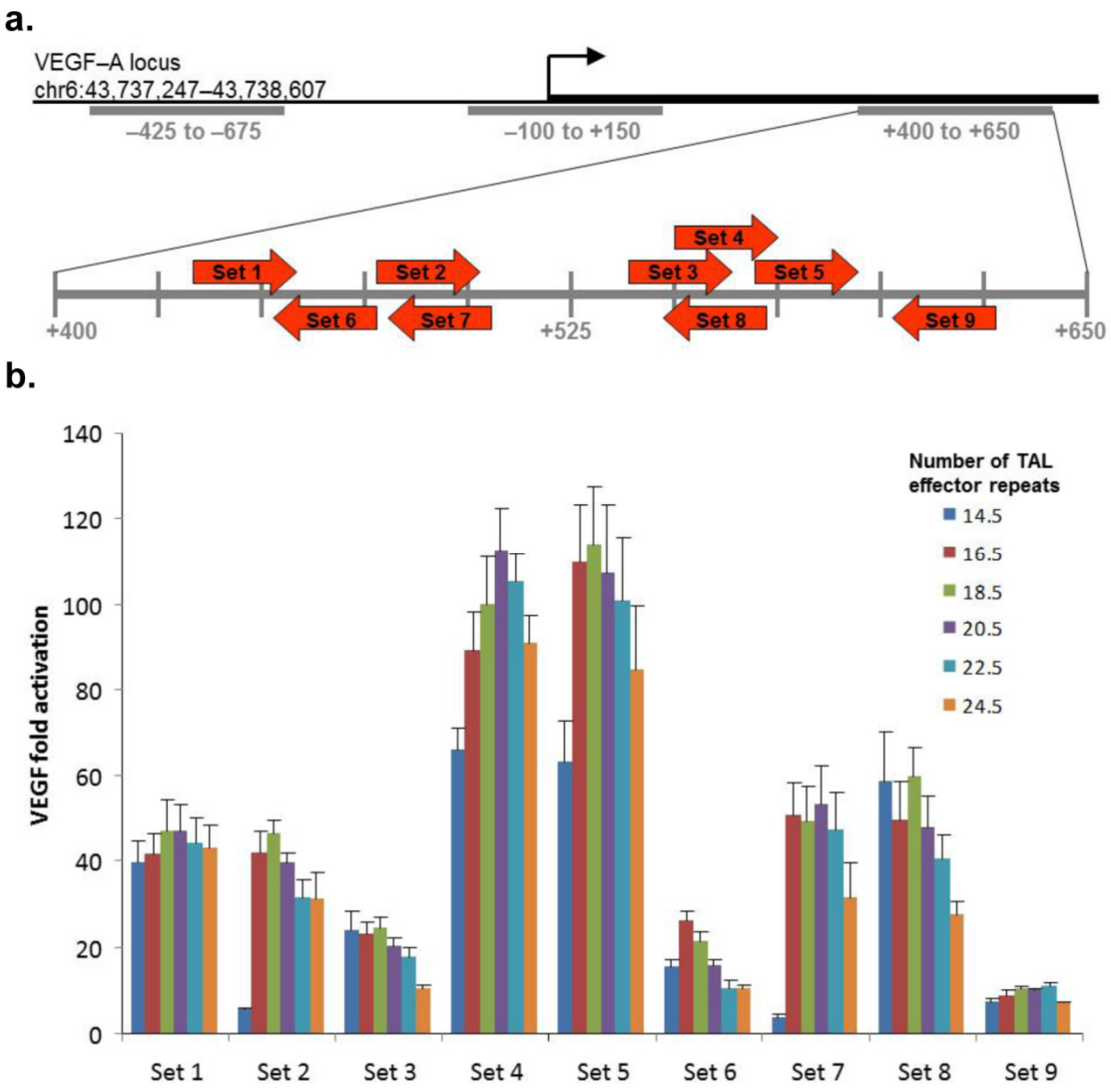


Figure 3.1. Activities of 54 variable length TALE-activators targeted to the endogenous human VEGF-A gene. (**a**) Schematic depicting the human *VEGF-A* promoter region. The transcription startpoint is indicated with a black arrow and previously published DNase I hypersensitive regions[13] are shown as grey bars. The DNase I hypersensitive region located between positions +400 to +650 relative to the transcription start site has been expanded, with red arrows indicating the locations and orientations of the 26 bp sites bound by the longest length TALE-activator (harboring 24.5 TAL effector repeats) in each set. (**b**) Activation of VEGF-A protein expression in 293 cells by 54 variable-length TALE-activators. Cells were transfected in triplicate with plasmids encoding each TALE-activator and then assayed as described in Methods. Mean fold-activation values are shown with error bars representing standard errors of the mean. All activators tested (except the 14.5-repeat activator from set 7) induced fold-activation of VEGF-A expression to a value significantly greater than 1, as determined by a one-sided, paired t-test.

transcription startpoint (Figure 3.1a) because previously published work with artificial zinc finger-based transcriptional activators has shown that targeting of sequences in HSSs greatly enhances success rates.[1] To investigate whether TALE-activators exhibit strand-dependent behavior, we targeted DNA sites that are on the sense or anti-sense strand of the *VEGF-A* gene (Figure 3.1). Using FLASH, we constructed sets of six variable-length TALE-activators (composed of 14.5, 16.5, 18.5, 20.5, 22.5, or 24.5 TAL effector repeats) for each of the nine target regions. All 54 of these proteins were built on a single common architecture similar to one previously described by Rebar and colleagues (and used to build the most highly active TALE-activator reported in the literature to date[4] (Methods and Supplementary Figure 3.1)) but that harbors a VP64 (instead of a VP16) activation domain. Strikingly, we found that 53 of these 54 TALE-activators induced significant increases in VEGF-A protein expression ranging from 5.3- to 114-fold (average of 44.3-fold activation) (Figure 3.1b) and that these activities do not appear to depend on which strand of DNA is bound. We do not know precisely why we observed variability in the range of fold-activations induced by these proteins but possibilities include variable protein expression levels or differences in the DNA-binding activities of the repeat arrays. Regardless of mechanism, our results suggest that TALE-activators can function efficiently when they are targeted to a DNase I HSS within a target gene promoter.

To further test the robustness of VP64 TALE-activators, we built more proteins targeted to six additional sites in the human *VEGF-A* promoter, to five sites in the human *NTF3* promoter, and to five sites in the microRNA *miR-302/367* cluster promoter. All of the TALE-activators we constructed were composed of 16.5 or 17.5 TAL effector repeats and were targeted to sites that lie within DNase I HSSs (Supplementary Figure 3.2 and Methods).

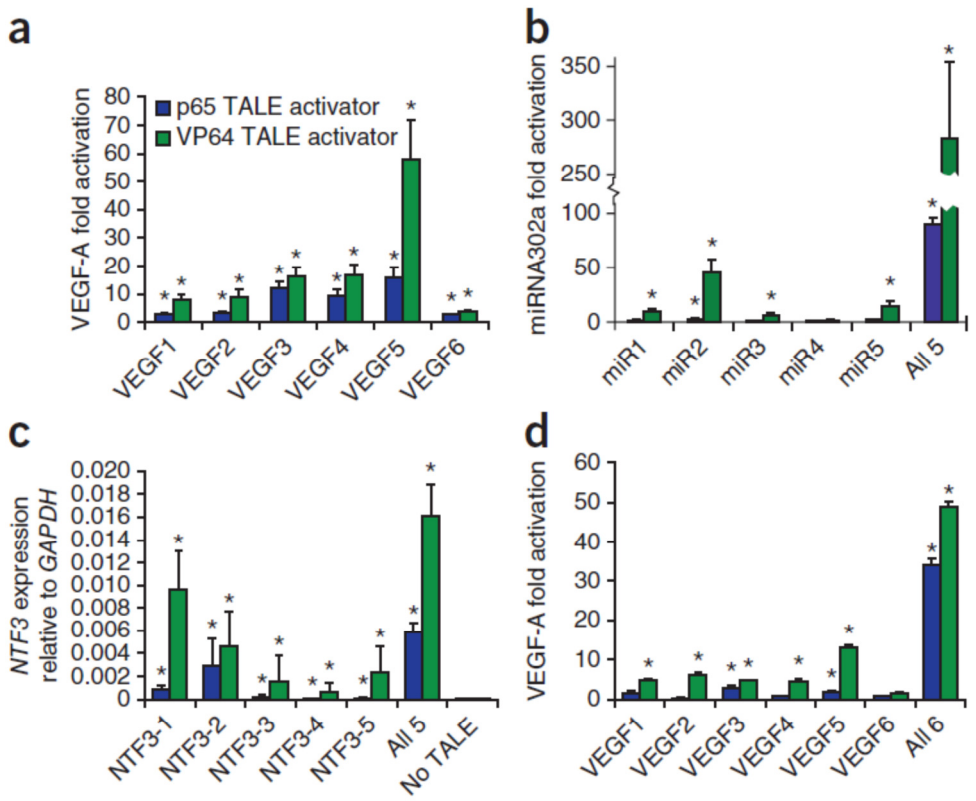


Figure 3.2. Activities of 16 TALE-activators targeted to the endogenous human VEGF-A, miR-302/367 cluster, and NTF3 genes. For all three gene targets, experiments were performed in triplicate with TALE-activators harboring either the VP64 (green bars) or NF-KB p65 (blue bars) activation domain. Error bars represent standard errors of the mean. (**a**) *VEGF-A*-targeted TALE-activators. Fold-activation values of VEGF-A protein were determined as described in Methods. Asterisks indicate activators that induced fold-activation of VEGF-A significantly greater than 1, as determined by a one-sided, paired t-test. (**b**) *miR-302/367*-targeted TALE-activators. Fold-activation values of *miR-302a* transcript were determined as described in Methods. Asterisks indicate activators that induced fold-activation of *miR-302a* transcript levels to a level significantly greater than 1 as determined by a one-sided, paired t-test. (**c**) *NTF3*-targeted TALE activators. Expression levels of *NTF3* mRNA relative to *GAPDH* mRNA are shown. Asterisks indicate activators that induced significant elevation of *NTF3* transcript levels relative to a control as determined by a one-sided, paired t-test. (**d**) Synergistic activation of VEGF-A. Experiments were performed as in (**a**) except that the amount of DNA used for individual TALE-activators was six-fold lower. Asterisks indicate activators that induced fold-activation of VEGF-A significantly greater than 1, as determined by a one-sided, paired t-test.

Notably, all six TALE-activators targeted to *VEGF-A* and four of five activators targeted to the *miR-302/367* cluster induced significant increases in the expression of their target genes in human transformed HEK293 and primary BJ fibroblasts, respectively (Figure 3.2a and 3.2b). Because *NTF3* mRNA is expressed at an essentially undetectable level in the HEK293 cells used for our experiments, we were unable to reliably quantify fold-activation values for proteins targeted to this gene, but all five TALE-activators significantly increased expression of *NTF3* relative to a *GAPDH* control (Figure 3.2c). Overall, 15 of these 16 TALE-activators (~94%) induced significant increases in expression of their endogenous gene targets (Figure 3.2a, 3.2b, and 3.2c, green bars).

We next explored whether replacing the VP64 domain in our TALE-activators with an NF-KB p65 domain would lead to consistently higher or lower levels of target gene expression. To do this, we substituted the VP64 activation domain with the NF-KB p65 domain for the 16 TALE-activators targeted to the *VEGF-A*, *NTF3*, and *miR302/367* promoters. For the 15 target sites for which we had obtained active TALE-activators, we found that the mean fold-activation induced was lower with the NF-KB p65 TALE-activators than with their matched VP64 counterparts (Figure 3.2a, 3.2b, and 3.2c, compare green and blue bars). We conclude that substitution with an NF-KB p65 domain can provide a general mechanism for reducing the fold-activation induced by a TALE-activator bearing a VP64 domain.

Because we were able to generate multiple active TALE-activators for each of the three genes we targeted, we also sought to examine whether several proteins working simultaneously on a single promoter could induce higher levels of fold-activation. Naturally occurring transcriptional activators can exhibit synergy – that is, the fold-activation observed in

the presence of multiple proteins is higher than the additive effects of the individual proteins.[14] Activator synergy enables both combinatorial and graded control of transcription in eukaryotes. We tested combinations of five VP64 or five p65 TALE-activators for their abilities to activate the *miR-302/267* cluster or the *NTF3* gene. Despite the fact that each TALE-activator plasmid in these combinations was present at one-fifth the level used when each was tested individually, we found that the use of multiple proteins led to substantially elevated transcription of the target genes (Figure 3.2b and 3.2c). Synergistic activation was observed with VP64 and p65 activators on the *miR-302/367* cluster (Figure 3.2b) and with p65 activators on the *NTF3* gene (Figure 3.2c). We also tested combinations of six VP64 and six p65 TALE-activators for their abilities to activate the *VEGF-A* gene but did not observe synergistic increases in expression (data not shown). We hypothesized that the lack of observed synergy might be due to some of these TALE-activators already maximizing transcription from the *VEGF-A* promoter. Consistent with this theory, we found that both the VP64 and p65 TALE-activators could mediate synergistic increases in VEGF-A expression under conditions where the same amounts of activator plasmids were used when tested individually or in combinations (Figure 3.2d). We conclude that both VP64 and p65 TALE-activators can function synergistically to increase expression of endogenous human genes and that this mechanism can be exploited to induce even higher levels of activation than can be achieved using only individual activators.

Discussion

Our ability to robustly construct highly active TALE-activators contrasts with the collective results of previous studies that described proteins with often much lower activities on

endogenous gene targets. We found that 62 of 65 (~95%) VP64 TALE-activators (for which we could calculate a fold-activation value) increased expression of their endogenous target gene by five-fold or more. By comparison, previously published studies found that only 4 of 26 (~15%) TALE-activators increased expression of their endogenous gene target by five-fold or more. We theorize that our targeting of sites in DNase I HSSs, as has been reported for engineered zinc finger transcription factors,[1] may be an important factor in our higher success rate because these regions not only represent open chromatin but may also encompass the binding sites for endogenous transcription factors. However, this hypothesis remains to be formally tested in future experiments. Additionally, as others have recently shown, the sequence architecture of the TALE repeats themselves (i.e.—the amino acids present at non-hypervariable positions within the repeats) may also influence DNA-binding activity.[15] In this regard, we note that all of the TALE repeat arrays tested in this study were constructed using our FLASH assembly method[13] on a single common architecture. Although there may be multiple potential reasons for this higher success rate, our results clearly demonstrate that highly active TALE-activators can be robustly constructed.

Our findings also expand the types of genes and the range of DNA sequences that can be targeted by TALE-activators. To our knowledge, our results provide the first demonstration that it is possible to activate a non-coding gene (*miR302/367*), thereby broadening the range of potential gene target types for TALE-activators. In addition, an analysis of our data suggests that there are no significant limitations in the range of sequences we can successfully target. Specifically, we found that restrictive targeting guidelines published in a previous report[16] need not be followed to obtain active TALE-activators (Supplementary Discussion). This resulting

capability to make TALE-activators for a broader range of DNA sequences is important because there are at least two factors that may potentially limit where one targets within a given gene: (1) a previous study[1] suggests that artificial transcriptional activators should ideally be designed to sites within DNase I HSSs and (2) the choice of the specific sequence targeted may help to minimize off-target effects (see below). Given this expanded targeting range of TALE-activators, we have updated our Zinc Finger and TALE (**ZiFiT**) Targeter software (http://zifit.partners.org) to assist potential users with identifying target sites in their genes of interest (Supplementary Discussion). Taken together, our results provide strong experimental support for the idea that TALE-activators can be used to control the expression of essentially any gene.

We have shown that TALE-activators can be used to regulate target genes across a wide dynamic range of expression, an important capability that will enable a broader range of applications for this technology. For example, for the three human genes we targeted in this report, we achieved 2.6- to 114-fold activation for VEGF-A protein expression, 2.8- to 283-fold activation of *miR302a* RNA, and elevated levels of *NTF-3* RNA that varied by 108-fold from lowest to highest. Our studies demonstrate three different approaches that can be used to fine-tune the level of gene expression induced by TALE-activators: (1) Varying the position of TALE-activator binding within a single DNase I HSS. We do not currently understand why targeting to different sites within a particular DNase I HSS leads to variable levels of activation (or why some TALE-activators fail to show activity). Our results at the *VEGF-A* locus do not suggest any obvious correlation between activity and distance, location, and/or orientation of the binding site relative to the transcription start site. Potential explanations include the

displacement of naturally occurring transcription factors and/or the methylation status of the target DNA site. Nonetheless, although it is currently not possible to predict the level of activation induced from any given site, the high success rate and broad targeting range of TALE-activators make it straightforward for an investigator to construct and empirically characterize a series of proteins to find one that induces the desired level of target gene expression; (2) Varying the transcriptional activation domain in a TALE-activator. For example, in the two human cell lines we examined, VP64 TALE-activators generally induced higher levels of gene expression than matched counterparts bearing a NF-KB p65 activation domain; and (3) Combining multiple TALE-activators can lead to even greater increases in target gene expression. Our finding that TALE-activators can synergistically activate transcription further broadens the range of gene expression changes that can be achieved with this platform. Perhaps more importantly, this finding raises the exciting possibility that target genes might be made responsive to multiple inputs, as has recently been shown with engineered zinc finger transcription factors.[17] The greater targeting range of engineered TALE-activators relative to artificial zinc finger activators provides a significant advantage for enabling synthetic biology applications in which artificial circuits are designed to interface with endogenous genes.

Although we have demonstrated that the FLASH platform can be used to construct highly active TALE-activators, future work will also need to address the specificities of such proteins. At present, relatively little is understood about the specificities of TALE-activators but we note that proteins harboring 16.5 or more repeats should bind to sequences of sufficient length (18 or more bps) to have a high probability of being unique in a complex genome. Our demonstration that the targeting range of TALE-activators is substantially greater than

previously suggested will also provide greater flexibility for choosing target sites to minimize potential off-target effects (once methods for determining these undesired binding events have been developed). In addition, two reports have recently suggested that a TAL effector repeat bearing hypervariable residues NH may be more specific for G than the repeat bearing NN (which in some contexts can also bind to A) that we used in all of our TALE-activators.[10, 15] One of these reports suggested that this NH repeat may bind more weakly than the NN repeat and therefore described recommendations (based on transient transfection reporter assays in plants) for how and when to use this NH repeat without compromising activity.[15] Construction and testing of variants based on our 68 successful VP64 activators could provide a large-scale prospective test of these guidelines and of the effect of substituting NN repeats with NH repeats on the activities and specificities of TALE-activators at their endogenous gene targets in human cells.

In this report, we have demonstrated that TALE-activators should function efficiently to regulate essentially any protein-coding or non-coding gene in human cells. This capability provides a useful complement to cDNA overexpression or RNAi-based regulation strategies for studying gene function and to previously described synthetic biology strategies that could potentially be used to regulate endogenous gene expression[18-22]. Our successes with TALE-activators in human cells should encourage use of these proteins in other cell types and organisms. More importantly, our findings should inspire the generation of other monomeric TAL effector-based fusion proteins. For example, TALE-repressors of transcription have been described recently[7, 10, 23] and fusions of engineered TAL effector repeat arrays to other monomeric functional domains (e.g.—enzymes that covalently modify DNA or histones) would

offer powerful new capabilities to rationally alter gene expression and/or epigenetic states. Thus, our findings should stimulate efforts to expand the repertoire of engineered TAL effector-based tools available for research, synthetic biology, and therapeutic applications.

Methods

Selection of TALE-activator binding sites

For the human *VEGF-A* gene, target sites were chosen that fall within DNase I hypersensitive sites previously described for 293 cells.[1] For the *NTF3* and *miR-302/367* cluster genes, target sites were chosen within digital DNase I hypersensitive regions identified from University of Washington ENCODE data using the UCSC genome browser (http://genome.ucsc.edu/);[24] we targeted these regions because they have been identified as DNase I hypersensitive sites in multiple different cell types and we therefore reasoned that these areas had a high probability of being in open chromatin.

Construction of TALE Activators

DNA fragments encoding TAL effector repeat arrays were generated using the Fast Ligation-based Automatable High-throughput Assembly (FLASH) method as previously described.[13] These fragments were cloned using overhangs generated by digestion with BsmBI restriction enzyme into expression vectors containing an EF1α promoter and the Δ152 N-terminal and +95 C-terminal TALE-derived domains from the previously described TALE-activator NT-L+95.[4] NF-KB p65 and VP64 activation domains were fused directly to the C-terminal end of the +95 domain and all fusion proteins harbor a nuclear localization signal.

Cell Culture and Transfection

Human Flp-In T-REx 293 cells and primary human BJ fibroblasts were maintained in Advanced DMEM supplemented with 10% FBS, 1% penicillin-streptomycin and 1% Glutamax (Life Technologies). Cells were transfected using either Lipofectamine LTX (Life Technologies) or Nucleofection (Lonza) according to manufacturer's instructions. Briefly, for experiments targeting *VEGF-A* and *NTF3* expression, 160,000 Flp-In T-REx 293 cells were seeded in 24-well plates and transfected the following day with 300 ng of plasmid encoding the TALE-activator (except in the reduced-concentration VEGF-A experiments (Figure 2d) in which 50 ng of plasmid encoding TALE-activator and 250 ng of a control plasmid expressing only the activation domain were transfected), 30 ng of pmaxGFP plasmid (Lonza), 0.5 µl Plus Reagent and 1.65 µl Lipofectamine LTX. For the synergy experiments with *VEGF-A* and *NTF3*, we transfected cells with 300 ng of TALE-activator-encoding plasmids (50 ng of each of six TALE-activator-encoding plasmids for *VEGF-A* and 60 ng of each of five TALE-activator-encoding plasmids for *NTF3*). For experiments targeting *miR-302/367* cluster expression, $5x10^5$ BJ fibroblasts were Nucleofected with 10 µg of plasmid encoding TALE-activator and 500 ng of pmaxGFP plasmid using the NHDF kit (Lonza) and program U-023 on the Nucleofector 2b device. For the synergy experiment with miR-302/367, we transfected cells with 10 µg of TALE-activator-encoding plasmids (2 µg of each of five TALE-activator-encoding plasmids).

ELISA Assays

Flp-In TREx 293 cells were transfected with plasmids encoding TALE-activators targeted to the human *VEGF-A* gene. All transfections were performed in triplicate. Cell media was harvested 40 hours after transfection and secreted VEGF-A protein levels in the media were assayed using a Human VEGF-A ELISA kit (R&D Systems). All samples were measured according to the manufacturer's instructions. Fold-activation values were calculated by dividing mean VEGF-A levels from media harvested from cells transfected with plasmids expressing TALE-activators by mean VEGF-A levels from cells transfected with plasmid expressing only the VP64 or p65 activation domain.

Quantitative RT-PCR assays

To measure *NTF3* mRNA levels, cells were harvested 2 days post-transfection and total RNA was isolated using the TRIzol Plus RNA purification system (Ambion). RNA was reverse transcribed using SuperScript III First-Strand Synthesis SuperMix and oligo-dT primer (Life Technologies). qPCR was then performed using the following Taqman primer/probe sets, as previously described[4] except with the modification that the GAPDH probe was labeled with HEX to allow for multiplexing - NTF3 forward primer: 5'-GATAAACACTGGAACTCTCAGTGCAA-3'; NTF3 reverse primer: 5'-GCCAGCCCACGAGTTTATTGT-3'; NTF3 taqman probe: 5'-/56-FAM/CAAACCTAC/ZEN/GTCCGAGCACTGACTTCAGA/3IABkFQ/-3'; GAPDH forward primer: 5'-CCATGTTCGTCATGGGTGTGA-3'; GAPDH reverse primer: 5'-CATGGACTGTGGTCATGAGT-3'; GAPDH taqman probe: 5'-/5HEX/TCCTGCACC /ZEN/ACCAACTGCTTAGCA/3IABkFQ/-3'. All TALE-activator-encoding plasmids and control plasmids were introduced into cells by Nucleofection in triplicate and qRT-PCR was performed in triplicate on each sample.

To measure *miR-302a* transcript levels, cells were harvested 3 days post-transfection and GFP-positive cells were isolated by flow cytometry. Total miRNA was isolated using the mirVana miRNA Isolation Kit (Ambion). Reverse transcription and qPCR were performed according to manufacturer's instructions using Applied Biosystems Taqman microRNA Assays (cat. #000529 for has-miR-302a and cat. #001006 for RNU48 control). Fold-activation of *miR-302a* RNA transcripts was calculated by comparing transcript levels from BJ fibroblasts transfected with plasmids encoding TALE-activators to transcript levels from BJ fibroblasts transfected with control plasmids expressing only the VP64 or p65 activation domains and using the comparative C_T ($\Delta\Delta C_T$) method. All TALE-activators and controls were introduced into cells by Nucleofection in triplicate and qRT-PCR for *miR302a* transcript and small RNA control *RNU48* were performed in triplicate on each sample.

Acknowledgements

This work was supported by a National Institutes of Health (NIH) Director's Pioneer Award DP1 OD006862 (J.K.J.), the Jim and Ann Orr MGH Research Scholar Award (J.K.J.), a National Science Foundation Graduate Research Fellowship (M.L.M.), and NIH T32 CA009216 (J.D.S.). We thank Jonathan E. Foley for technical assistance with construction of TALE-activator plasmids, Ravi Mylvaganam for technical assistance with flow cytometry, and the Massachusetts General Hospital (MGH) Nucleic Acid Quantitation Core (supported by NIH P30 NS45776) for assistance with performing real-time RT-PCR assays.

Conflict of interest statement

J.K.J. has a financial interest in Transposagen Biopharmaceuticals. J.K.J.'s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

References

1. Liu, P.Q. et al. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem* **276**, 11323-11334 (2001).

2. Joung, J.K. & Sander, J.D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **advance online publication** (2012).

3. Mussolino, C. & Cathomen, T. TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol* (2012).

4. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148 (2011).

5. Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153 (2011).

6. Geissler, R. et al. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One* **6**, e19509 (2011).

7. Garg, A., Lohmueller, J.J., Silver, P.A. & Armel, T.Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* (2012).

8. Tremblay, J.P., Chapdelaine, P., Coulombe, Z. & Rousseau, J. TALE proteins induced the expression of the frataxin gene. *Hum Gene Ther* (2012).

9. Wang, Z. et al. An Integrated Chip for the High-Throughput Synthesis of Transcription Activator-like Effectors. *Angew Chem Int Ed Engl* **51**, 8505-8508 (2012).

10. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**, 968 (2012).

11. Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* **40**, 5368-5377 (2012).

12. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).

13. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**, 460-465 (2012).

14. Carey, M., Lin, Y.S., Green, M.R. & Ptashne, M. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* **345**, 361-364 (1990).

15. Streubel, J., Blucher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat Biotechnol* **30**, 593-595 (2012).

16. Doyle, E.L. et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40**, W117-122 (2012).

17. Khalil, A.S. et al. A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012).

18. Deans, T.L., Cantor, C.R. & Collins, J.J. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363-372 (2007).

19. Tigges, M., Marquez-Lago, T.T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309-312 (2009).

20. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333**, 1307-1311 (2011).

21. Culler, S.J., Hoff, K.G. & Smolke, C.D. Reprogramming Cellular Behavior with RNA Controllers Responsive to Endogenous Proteins. *Science* **330**, 1251-1255 (2010).

22. Auslander, S., Auslander, D., Muller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* **487**, 123-127 (2012).

23. Mahfouz, M.M. et al. Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol* **78**, 311-321 (2012).

24. Rosenbloom, K.R. et al. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res* **40**, D912-917 (2012).

Chapter 4

Targeted DNA Demethylation and Endogenous Gene Activation with Programmable TALE-TET1 Proteins

Morgan L. Maeder[1,2,3], James F. Angstman[1,2], Marcy E. Richardson[4], Samantha J. Linder[1,2], Vincent Cascio[1,2], Shengdar Q. Tsai[1,2,5], Quan Ho[1,2], Deepak Reyon[1,2,5], Jeffry D. Sander[1,2,5], Joseph F. Costello[6], Miles Wilkinson[4], J. Keith Joung[1,2,3,5,*]

[1]Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129 USA

[2]Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA 02129 USA

[3]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02115 USA

[4]Department of Reproductive Medicine, University of California San Diego, La Jolla CA 92093

[5]Department of Pathology, Harvard Medical School, Boston, MA 02115

[6]Brain Tumor Research Center, University of California San Francisco, San Francisco, CA 94143

This manuscript is in preparation for submission to *Nature Biotechnology*.

Morgan Maeder designed and oversaw experiments. Morgan Maeder designed TET1 fusion constructs, oversaw construction of these plasmids by James Angstman, performed HBB and RHOXF2 transfections, qRT-PCR for HBB experiments and analyzed all qRT-PCR and high-throughput bisulfite sequencing data. James Angstman performed KLF4 transfections and bisulfite sequencing library prep, Samantha Linder and Vincent Cascio performed RHOXF2 and HBB bisulfite sequencing library prep, Marcy Richardson performed RHOXF2 qRT-PCR, Shengdar Tsai performed de-multiplexing and initial analysis of MiSeq data, Quan Ho performed Western blots and Jeffry Sander engineered zinc fingers.

Abstract

Recent large-scale, genome-wide studies have defined cell-type specific patterns of DNA methylation, a critical mechanism used to regulate gene expression in both normal development and disease states. However, determining the functional significance of specific methylation marks remains a challenging problem due to the lack of targeted methodologies for perturbing these modifications. Here we describe an approach for efficient targeted demethylation of specific CpGs in human cells using fusions of engineered transcription activator-like effector (TALE) repeat arrays and the TET1 hydroxymethylase catalytic domain. Using these TALE-TET1 reagents, we demonstrate that loss of methylation at critical promoter CpG positions can trigger substantial increases in endogenous gene expression in human cells. Our results delineate a general strategy for probing the functional significance of specific CpG methylation marks in the context of endogenous genes and validate new programmable DNA demethylation reagents with broad utility for research and potential therapeutic applications.

Introduction

Methylation of DNA cytosine is an important mechanism widely used to regulate patterns of gene expression in higher eukaryotic organisms[1]. In mammalian cells, the generation of methylated cytosine (5mC) is catalyzed and maintained by DNA methyltransferases (DNMTs) primarily at CpG dinucleotides[2]. Demethylation of 5mC is believed to be mediated by the ten-eleven translocation (TET) family of proteins, enzymes that catalyze the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), a critical step that ultimately leads to removal of the methyl mark[3-7]. A substantial literature, in particular recent work from the

ENCODE project, has shown that regions of hypermethylated DNA in mammalian cells can be associated with silenced, inactive chromatin whereas regions of hypomethylated DNA can be associated with expressed genes and open chromatin[8,9]. DNA methylation has been shown to play a critical role in determining tissue-specific expression of genes during cell-type differentiation and normal development[10]. The hypermethylation of DNA has also been linked with various human diseases, in particular with multiple types of cancer including metastatic melanoma, renal carcinoma and breast and prostate cancer[11-15].

Although DNA hypermethylation has been associated with gene silencing and various disease states, demonstrating and understanding the causal effects of specific CpG methylation events has remained challenging due to the lack of targeted methods for removing methyl marks. At present, only non-specific approaches exist for reducing or removing methyl groups from CpGs. For example, the small molecule 5-azacytidine (5-aza), an inhibitor of DNMTs, has been widely used to study the effects of demethylation on specific gene promoters. However, 5-aza leads to global demethylation of CpGs in cells, making it difficult to definitively establish causal effects. To our knowledge, no method currently exists that enables researchers to specifically demethylate CpGs in a targeted fashion. Here we sought to develop this capability by fusing the hydroxymethylase activity of TET1 to engineered transcription activator-like effector (TALE) repeat arrays with programmable DNA-binding specificities. Customized TALE repeat arrays make an attractive platform for directing TET1 activity because monomeric proteins that bind to nearly any target DNA sequence of interest can be robustly made by assembling together individual repeat domains with known single base specificities[16,17].

Results

In initial experiments, we sought to define the architecture of a TALE-TET1 fusion protein that could mediate targeted demethylation of CpGs. To do this, we fused TALE repeat arrays engineered to bind two sites in the human *KLF4* gene to either full-length human TET1 or its catalytic domain (CD) (Figures 4.1a and 4.1b; Methods) and tested the abilities of these proteins to mediate demethylation of adjacent CpGs in human K562 cells. For both target sites, TALE fusions bearing the TET1 CD domain induced significantly greater decreases in methylation of CpGs proximal to the TALE binding than those bearing the full-length TET1 protein (Figures 4.1c and 4.1d). For example, robust demethylation was observed in cells treated with the TALE-TET1CD fusion KLF-1, which reduced the methylation at CpGs located 10 and 16 bp from the 3' boundary of the TALE binding site by 54% and 66%, respectively. Interestingly, this KLF-1 fusion also induced demethylation of CpGs located ~150 bp away from 3' edge of its binding site, albeit to a lesser degree than the more proximal CpGs (Figure 4.1c). Control fusion proteins bearing a TALE repeat array targeted to an unrelated EGFP reporter gene sequence did not demethylate CpGs in the *KLF4* intron (Figures 4.1c and 4.1d), demonstrating that this activity requires specific binding to the target locus by the TALE repeats and is not due simply to overexpression of proteins harboring TET1 hydroxymethylase activity. Lengthening the linker between the TALE repeat array and the TET1 CD did not appreciably alter demethylase activity (Supplementary Figure 4.1). Based on these results, all additional fusion proteins described in this report were constructed on the TALE-TET1CD architecture with a short GGGS linker (hereafter referred to as simply "TALE-TET1" proteins).

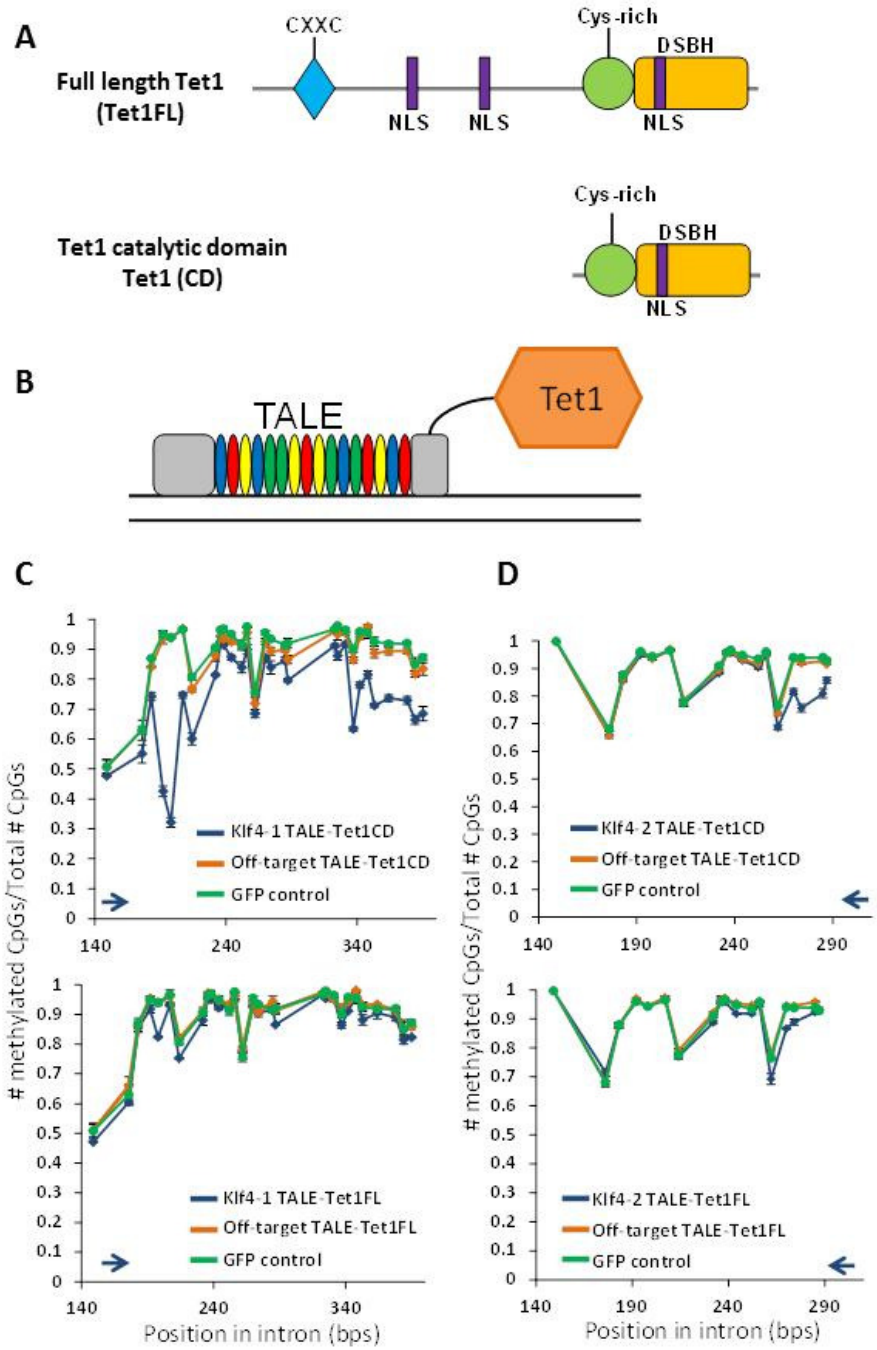


Figure 4.1. TALEs fused to Tet1FL and Tet1CD A) Schematic shows the predicted domain architecture of the Tet1 protein and the catalytic domain of the Tet1 protein. Features shown include the CXXC-type zinc-binding domain (CXXC), three nuclear localization signals (NLS), the cysteine-rich region (cys-rich) and the double-stranded β helix domain (DSBH) (Tahiliani et al., *Science* 2009). **B)** Schematic shows TALE-Tet1 fusion protein. **C)** Methylation status of CpGs in Klf4 intron 1/2 in K562 cells transfected with KLF4 -1 (blue) or an off-target TALE (orange) fused to either Tet1CD (top) or Tet1FL (bottom) or a GFP control (green). Blue arrow shows the location of TALE binding site. Error bars represent sem of three independent samples. Note that the GFP control is the same for both graphs and is depicted twice for ease of comparison. **D)** Same as in C but with KLF4 TALE-2.

We next determined whether TALE-TET1-induced demethylation of specific CpGs in human promoters might induce changes in proximal endogenous gene expression. The *RHOXF2* gene is expressed primarily in germ cells[18] and reporter expression studies have suggested that expression may be controlled by the methylation status of a small number of CpGs in the promoter (M. Richardson and M. Wilkinson, unpublished data). We engineered eleven TALE-TET1 proteins targeted to sites that lie in close proximity to a total of 18 different CpGs in the *RHOXF2* promoter (Figure 4.2a). Based on initial screening using bisulfite sequencing and quantitative RT-PCR (data not shown), we identified two TALE-TET1 proteins (RHOX-3 and -4) that demethylated proximal CpGs in the -200 to +1 region of the *RHOXF2* promoter and also induced significant increases in *RHOXF2* mRNA expression in two different human cell lines (293 and HeLa; Figures 4.2b-2e). RHOX-3 can also bind to an additional site in the -650 to -750 region of the RHOXF2 promoter but demethylation of CpGs in this region is only observed in 293 cells because cytosines in this region are not methylated in HeLa cells (Supplementary Figure 4.2). Interestingly, we found that even greater increases in RHOXF2 expression could be induced by combined expression of both the RHOX-3 and -4 TALE-TET1 proteins, particularly in 293 cells (Figures 4.2c and 4.2e). The higher degree of activation observed in 293s may be due demethylation of an additional cluster of CpGs by RHOX-3 in those cells.

Although our results suggest that the TALE-TET1-induced increases in *RHOXF2* expression are caused by targeted promoter CpG demethylation, we wished to rule out potential alternative hypotheses such as competitive binding of TALE-TET1 fusions with endogenous transcription factors or the presence of a fortuitous transcriptional activation function within the fusion protein. To do this, we tested RHOX-3 and -4 proteins bearing

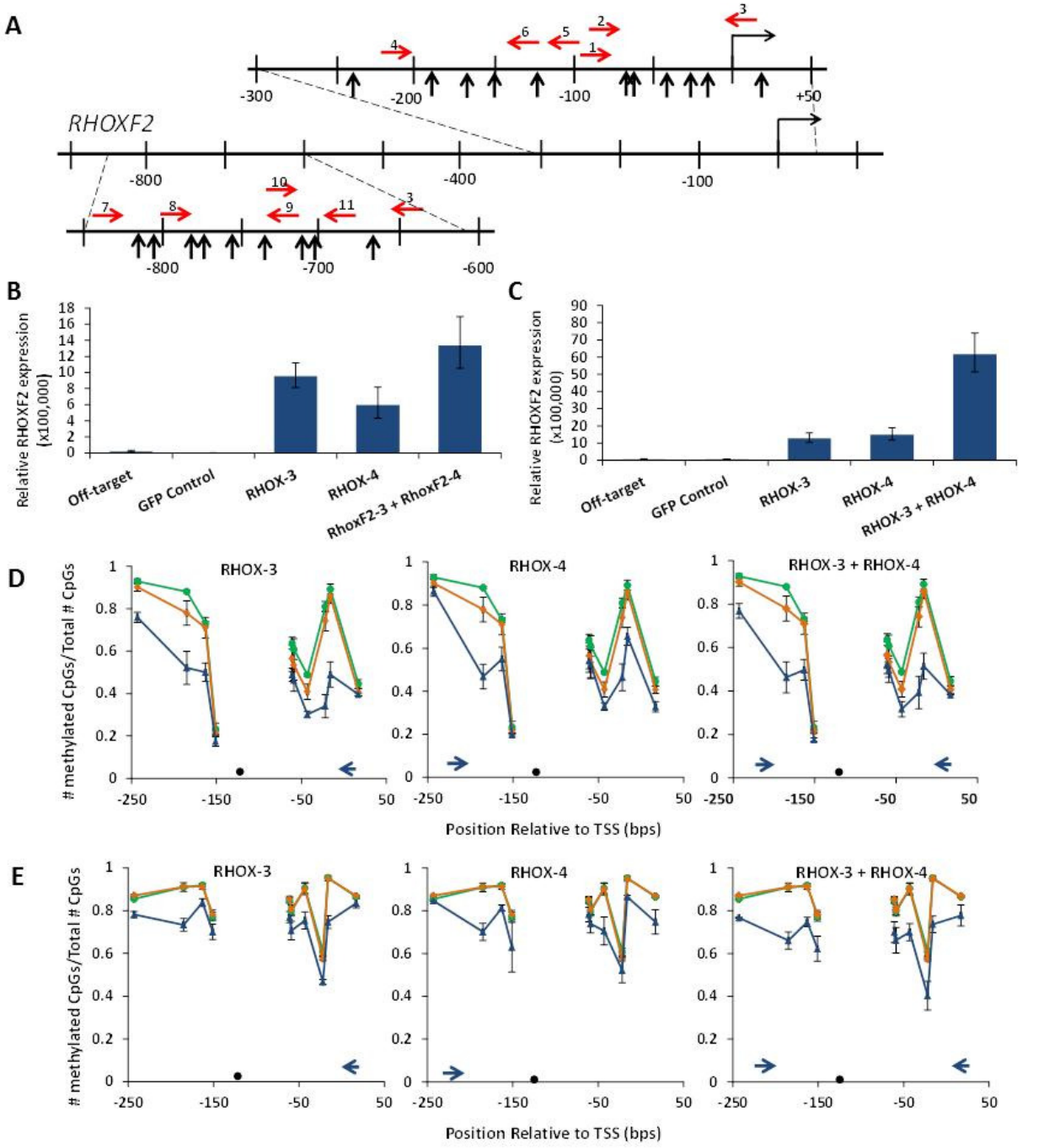


Figure 4.2. Demethylation of the Human RHOXF2 Locus. A) Schematic shows the RHOXF2 locus with CpGs indicated with black arrows. TALE binding sites are depicted as red arrows. **B)** Expression levels of RHOXF2 mRNA, relative to β-Actin, in HeLa cells as determined by qRT-PCR. Error bars represent sem of three independent samples, each assayed three times. All samples are statistically significant compared to the off-target control as determined by a one-sided T-test ($P<0.05$). **C)** Same as B but in Hek293 cells. **D)**Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in HeLa cells transfected with TALE-Tet1CD(s) targeted to the RHOXF2 locus (blue), an off-target TALE-Tet1CD control (orange) or a GFP control (green). Blue arrow depicts position of the TALE binding site. Black dot represents CpG at -125 that was not bisulfite sequenced due to limits of sequencing read lengths. Error bars represent sem of three independent samples. Note that the GFP and off-target controls are the same in all graphs. **E)** Same as C but in Hek293 cells.

mutations (H1671Y, D1673A) known to inactivate TET1 catalytic activity[3] and found these proteins neither demethylated their proximal CpGs nor activated RHOXF2 gene expression in both 293 and HeLa cells (Supplementary Figure 4.3). Western blots also confirmed that the observed inactivity of these mutated RHOX-3 and -4 proteins is not due to their decreased expression in 293 cells (Supplementary Figure 4.4). We conclude that specific demethylation of certain CpGs in the *RHOXF2* promoter can trigger increased gene expression.

To extend and generalize our finding that TALE-TET1 fusion proteins can identify CpG methylation marks that influence endogenous gene expression, we sought to demethylate CpGs in an additional locus, the human beta-globin (*HBB*) gene promoter. Previous work suggested that four CpGs, which are differentially methylated in erythroid cells isolated from fetal liver and adult bone marrow[19], may play a role in regulating *HBB* gene expression. To test this hypothesis, we constructed ten TALE-TET1 proteins targeted to various sites proximal to these four CpGs (Figure4. 3a). Although all ten TALE-TET1 fusions induced significant demethylation of CpGs near their respective binding sites in human K562 cells (Figure 4.3b and 4.3c), significant increases in HBB gene expression as measured by qRT-PCR were observed with only four of these proteins (HBB-3, -4, -5, and -6) (Figure 4.3d). Of note, the three proteins (HBB-4, -5, and -6) that induced the greatest fold-activation of the promoter were the only fusions that induced significant demethylation of the CpG at position -266 (numbered relative to the transcription start site) and the magnitude of HBB gene activation observed correlated with the extent of demethylation (Figure 4.3d). HBB-4, -5, and -6 proteins bearing the H1671Y/D1673A mutations that inactivate TET1 catalytic domain activity failed to demethylate the -266 CpG and also failed to efficiently activate HBB gene expression in K562 cells

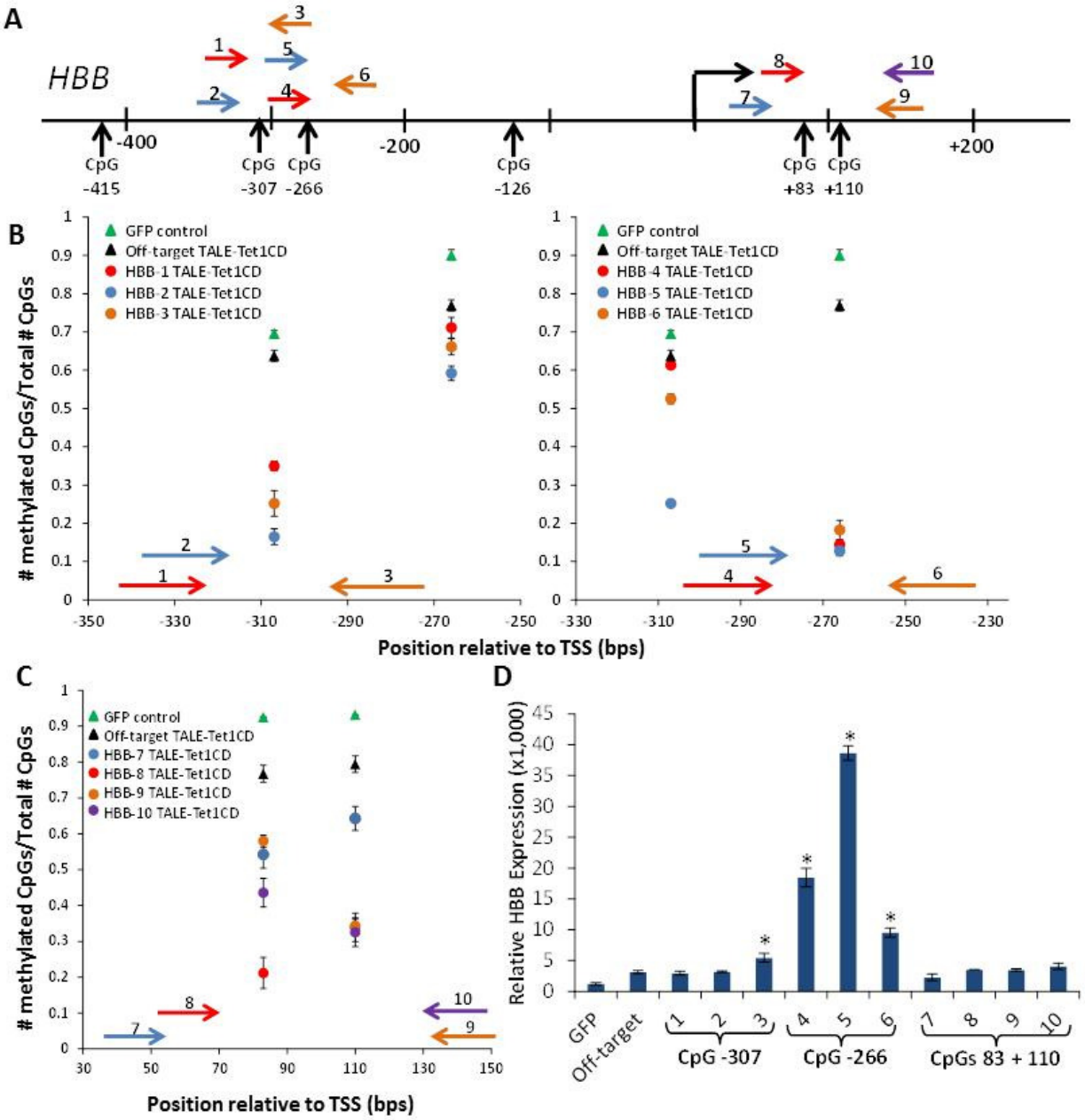


Figure 4.3. Demethylation of the Human β-globin Locus. A) Schematic shows the HBB locus with CpGs indicated with black arrows and labeled relative to the transcription start site. TALE binding sites are depicted as red arrows. **B)** Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in K562 cells transfected with TALE-Tet1CDs targeted to the HBB locus (red, orange and blue circles), an off-target TALE-Tet1CD control (black triangle) and a GFP control (green triangle). Position of TALE binding sites are depicted with colored arrows corresponding to the color of the circles on the graph. TALEs targeted to the -310 CpG are depicted in the left panel and TALEs targeted to the -266 CpG are dipicted on the right. Error bars represent sem of three independent samples. Note that the GFP and off-target controls are the same in both graphs and are depicted twice for ease of comparison. **C)** Same as in B but showing demethylation of the 2 CpGs located at +83 and +110. **D)** Expression levels of HBB mRNA, relative to β-Actin, as determined by qRT-PCR. Error bars represent sem of three independent samples and asterisks indicate statistical significance compared to the off-target control as determined by a one-sided T-test ($P<0.05$).

(Supplementary Figure 4.5a and 4.5b). Western blot experiments with HBB-5 and -6 confirm that the loss of demethylation and gene activation activities observed with these catalytically inactive mutants is not due to decreased protein expression in K562 cells (Supplementary Figure 4.5c-4.5e).

To test whether the -266 CpG demethylation and increased *HBB* gene expression induced by HBB-4, -5-, and -6 are stable over time, we assessed these variables in a more extended time-course experiment. K562 cells transfected with expression plasmids encoding HBB-4, HBB-5, HBB-6, or a control TALE-TET1 targeted to an EGFP sequence were assessed for -266 CpG demethylation and HBB gene expression at 4, 7, 14, and 30 days post-transfection. HBB-4, -5, and -6 all showed their highest levels of fold-activation at post-transfection day 4 with HBB expression steadily decreasing by day 30 to the same level observed with the control TALE-TET1 targeted to EGFP (Supplementary Figure 4.6a). Strikingly, for all three HBB-targeted TALE-TET1 proteins, the extent of -266 CpG demethylation paralleled the fold-activation of HBB gene expression while the methylation status of another CpG at position -306 did not change significantly over the time-course of this experiment (Supplementary Figure 4.6b). Taken together, our findings strongly suggest that the demethylation of the -266 CpG requires continued expression of a targeted TALE-TET1 protein and that demethylation of this particular CpG is critically required for the observed activation of HBB gene expression.

Discussion

Our results define, to our knowledge, the first generalizable approach for targeting DNA hydroxymethylase activity and subsequent cytosine demethylation to any genomic locus of

interest in living cells. The TALE-TET1 framework described here can be easily programmed to bind to essentially any DNA sequence using the simple TALE repeat code and we used this platform to induce efficient targeted demethylation at three endogenous genes (KLF4, RHOXF2, and HBB) in three different human cell lines. The demethylation of neighboring CpGs induced by these proteins is generally very focal, occurring within 30 nts of either end of the target binding site. However, with some proteins a cluster of demethylated CpGs 150 bps away from the target site could sometimes also be observed, suggesting that the TET1 CD can also access regions of open chromatin located one nucleosome distance away. More than half of the TALE-TET1 proteins we made showed demethylase activities but we observed variable success rates of 100% (2 out of 2), ~18% (2 out of 11), and 100% (10 out of 10) for the KLF4, RHOXF2, and HBB genes, respectively; therefore, it will be of interest in future experiments to define the locus-specific parameters (e.g.—chromatin structure, nucleosome occupancy) that might influence TALE-TET1 protein activities and success rates.

Our experiments provide a framework for evaluating the functional significance of specific CpG methylation events using TALE-TET1 proteins. Using this approach, we successfully defined several CpGs within the RHOXF2 promoter and a single CpG within the HBB promoter for which specific removal of methyl groups leads to an increase in gene expression. For the HBB promoter, additional support for functional significance was provided by the demonstration that progressive re-methylation of the critical CpG (which, not surprisingly, occurs as expression of the TALE-TET1 protein from transiently transfected plasmids is lost in cells over time) paralleled decreases in HBB gene expression. Although we acknowledge that the CpGs we have identified in our transformed cancer cell-based experiments may not

necessarily be relevant to normal physiologic regulation of the RHOXF2 and HBB genes, our proof-of-principle experiments illustrate how TALE-TET1 proteins can be used to define critical CpGs in other cell types with greater relevance for normal development or disease. Our TALE-TET1 tools should also facilitate more detailed mechanistic studies that define how the loss of methylation marks leads to increases in promoter activity.

An important and as-yet unanswered issue for future studies will be to define the genome-wide specificities of our TALE-TET1 proteins. All of the proteins we constructed for the RHOXF2 and HBB promoters were designed to bind 20 bp sites, sequences sufficiently long enough to be potentially unique in the human genome. However, although previously published *in vitro* SELEX experiments suggest that TALE repeat arrays are specific for their intended target sites, to our knowledge the genome-wide specificities of monomeric TALE proteins in human cells have not been previously described. Engineered zinc finger (ZF) proteins provide a potential alternative to TALE repeat arrays for targeting TET1 activity and at least one published report has suggested that monomeric six-finger proteins can be highly specific in human cells[20]. We also engineered multiple six-finger ZF-TET1 fusion proteins targeted to 18 bp sequences in the *KLF4* and *HBB* genes (Supplementary Methods) and demonstrated that these can induce targeted demethylation with efficiencies comparable to those induced by TALE-TET1 proteins (Supplementary Results and Supplementary Figures 4.7 and 4.8). Regardless of which platform is ultimately proven to be more specific, potential off-target effects can be readily controlled for by constructing and testing multiple targeted TALE-TET1 or ZF-TET1 fusion proteins for each CpG or cluster of CpGs to be demethylated. For example, our finding that three different TALE-TET1 proteins can all demethylate a common

CpG and induce changes in HBB gene expression strongly suggests that the observed common effect is due to binding at the intended target sequence and not at an off-target site elsewhere in the genome. We note that the greater targeting range of, and simplicity of engineering, TALE repeat arrays make them clearly preferable to ZFs for making multiple proteins.

The TALE-TET1 platform described here and the targeted histone H3K4 demethylase TALE-LSD1 platform (described in an accompanying manuscript from Bernstein and colleagues) represent novel and important additions to the growing toolbox of reagents for performing targeted editing of epigenomic modifications. These TET1- and LSD1-based proteins are the first programmable epigenome editing tools to utilize engineered TALE technology; by contrast, previously described reagents target histone methyltransferases (SUV39H1 and G9A[21]) or DNA methyltransferases (bacterial enzymes[22-26] and human DNMT3a and 3b subunits[27-29]) using engineered zinc finger proteins, which can be more challenging to construct and have a more restricted targeting range. More importantly, among these various targeted epigenome editing tools, our TALE-TET1 platform is the first to enable activation, as opposed to repression, of endogenous gene promoter activity. This raises the longer-term prospect of using transient expression of TALE-TET1 proteins together with other targeted epigenome modifier proteins to induce stable epigenetic activation of any poised or silenced gene of interest. Development of such a capability would enable numerous research applications as well as potential therapeutic strategies for epigenetic-based diseases.

Methods

TALE-TET1 Fusion Protein Design and Construction

The full TET1 coding sequence was synthesized (Integrated DNA Technologies) and used to construct TET1-FL and TET1-CD expression vectors. All TALEs were assembled using the FLASH method and were cloned into TALE-TET1 expression vectors containing an N-terminal nuclear localization signal, the Δ152 TALE N-terminal domain and the +95 TALE C-terminal domain as previously described (Miller et al., Nature Biotech 2011). TET1-FL or TET1-CD was fused C-terminal to the TALE with a GlyGlyGlySer linker and expression was driven by an EF1alpha promoter. For Western blot experiments, a triple-flag tag was cloned upstream of the nuclear localization signal.

ZF-TET1 Fusion Protein Design and Construction

Six-finger zinc finger arrays were assembled from two-finger units derived from OPEN-selected three-finger proteins. Linker sequences. ZFs were cloned into ZF-TET1 expression vectors containing an N-terminal nuclear localization signal. TET1-FL or TET1-CD was fused C-terminal to the ZF with a LeuArgGlySer linker and expression was driven by an EF1alpha promoter.

Cell culture and Transfection.

FlpIn-TRex HEK293 and HeLa cells were cultured in Advanced DMEM supplemented with 10% FBS, 1% Glutamax, and 1% penicillin-streptomycin (Invitrogen). K562 cells were maintained in RPMI supplemented with 10% FBS, 1% Glutamax, and 1% penicillin-streptomycin. TALE-Tet1

fusions targeted to the KLF4 and HBB loci were transfected into K562 cells using nucleofection. Briefly, 10 ug of TALE-Tet1 fusion plasmid and 500 ng pmaxGFP plasmid were nucleofected into $1x10^6$ K562 cells using Kit V (Lonza) and program T-016. GFP control transfections contain 500ng pma GFP plasmid only. TALE-Tet1 fusions targeted to human RhoxF2 were transfected into HEK293 and Hela cells using Lipofectamine LTX according to manufacturer's instructions (Invitrogen). Briefly, 320,000 293 cells or 100,000 Hela cells were seeded into 12 well plates and transfected the following day with 1.2 ug TALE-TET1 plasmid, 60 ng pmaxGFP plasmid, 1 ul Plus reagent, and 3.3 ul Lipofectamine LTX.

Genomic DNA and Total RNA isolation.

Four days post-transfection, genomic DNA (gDNA) was isolated using the QIAamp DNA Blood Mini Kit (Qiagen) according to manufacturer's protocol. Total RNA was isolated from cells transfected with plasmids encoding TALE-TET1 fusions targeting HBB and RhoxF2 using the PureLink RNA Mini Kit (Ambion) according to manufacturer's instructions. RNA was treated with TurboDNA-Free (Ambion).

High-Throughput Bisulfite Sequencing.

500 ng of genomic DNA isolated from transfected cells was bisulfite treated using the EZ DNA methylation, EZ DNA Methylation-Gold or EZ DNA Methylation-Lightning Kits (Zymo Research) according to manufacturer's protocol. Targeted KLF4, HBB, and RHOXF2 sites were amplified using bisulfite-converted gDNA as a template with Kapa Biosystems' Kapa HiFi HotStart Uracil+ ReadyMix (for KLF4 and HBB PCRs) or Qiagen's PyroMark PCR Kit (for RhoxF2

PCRs). Standard Illumina adaptors were added by either ligation or PCR and Illumina multiplex barcodes were added by PCR. For details of PCR reactions, see supplemental methods. Pooled amplicons were sequenced on an Illumina MiSeq using 150 bp paired-end reads (Dana Farber MBCF Genomics Core).

qRT-PCR assays

For HBB assays, RNA was reverse transcribed using the SuperScript III First-Strand Synthesis SuperMix and oligo-dT (Invitrogen). qPCR was performed with Taqman Universal PCR Mastermix (Applied Biosystems) on an ABI 7500 Fast Real-Time PCR system with the following primer/probe sets: Forward HBB primer 5'-CAAGGGCACCTTTGCCACAC-3'; Reverse HBB Primer 5'- TTTGCCAAAGTGATGGGCCA-3'; HBB Taqman Probe 5'-/56-FAM/CCTGGGCAA/ZEN/CGTGCTGGTCTGTGT/3IABkFQ/-3'; Forward ACTB Primer: 5'-GGCACCCAGCACAATGAAG-3'; Reverse ACTB primer 5'- GCCGATCCACACGGAGTACT-3'; ACTB Taqman Probe 5'-/5MAX550-Y/TCAAGATCA/ZEN/TTGCTCCTCCTGAGCGC/3IABlk_FQ/-3'.

For RHOXF2 assays, RNA was reverse transcribed using iScript cDNA Synthesis Kit (BioRad) according to manufacturer's protocol. qPCR was performed with SsoAdvanced SYBRGreen Supermix (BioRad) on an ABI StepOnePlus instrument with the following primers: RHOXF2 Forward primer 5'- GGCAAGAAGCATGAATGTGA-3'; RHOXF2 Reverse primer 5'-TGTCTCCTCCATTTGGCTCT-3'; M/H Actin Forward primer 5'- GTCCACACCCRCCGCCAG-3'; M/HActin Reverse primer 5'- CCCACGATGGAGGGGAA-3'.

Western Blotting

Cells were lysed in radioimmunoprecipitation assay protein extraction buffer (Cell Signaling) and centrifuged for 5 min at 4°C to harvest the supernatant. The concentration of the extracted protein was determined using Thermo Scientific protein assay kit (Thermo Scientific). Ten micrograms per lane of the extracted protein were loaded onto 4% to 12% Tris-glycine gels (Invitrogen) for electrophoresis, and electro-transferred to Nitrocellulose Membrane (Invitrogen). The membrane was blocked with 5% nonfat dry milk in Tris-buffered Saline-Tween buffer [25 mmol/L Tris (pH 7.4), 137 mmol/L NaCl, 2.7 mmol/L KCl, 0.1% Tween 20] for 1 hr at room temperature, followed by incubation with FLAG monoclonal antibody (Sigma) and β-actin (C4) HRP monoclonal antibody (Santa Cruz Biotechnology) in blocking buffer for 2 hr at room temperature. After washing with TBS-T buffer, the membrane was incubated with goat anti-mouse IgG (H+L) - HRP conjugate antibody (Bio-Rad Laboratories) at a dilution of 1:7500 in blocking buffer for 1 h at room temperature. The membrane was developed using ImmunoCruz chemiluminescence system (Santa Cruz Biotechnology) and exposed to BioMax XAR film (Kodak).

References

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16,** 6–21 (2002).
2. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9,** 2395–2402 (2000).
3. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
4. Guo, J. U., Su, Y., Zhong, C., Ming, G.-L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).
5. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333,** 1300–1303 (2011).
6. He, Y. F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in

Mammalian DNA. *Science* **333,** 1303–1307 (2011).
7. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42,** 451–464 (2011).
8. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
9. ENCODE Project Consortium *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
10. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–220 (2013).
11. Costello, J. F. *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24,** 132–138 (2000).
12. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat. Rev. Cancer* **4,** 143–153 (2004).
13. Soengas, M. S. *et al.* Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature* **409,** 207–211 (2001).
14. Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci USA* **91,** 9700–9704 (1994).
15. Graff, J. R. *et al.* E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Res.* **55,** 5195–5199 (1995).
16. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326,** 1509–1512 (2009).
17. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* (2012). doi:10.1038/nbt.2170
18. Wayne, C. M., MacLean, J. A., Cornwall, G. & Wilkinson, M. F. Two novel human X-linked homeobox genes, hPEPP1 and hPEPP2, selectively expressed in the testis. *Gene* **301,** 1–11 (2002).
19. Mabaera, R. *et al.* Developmental- and differentiation-specific patterns of human gamma- and beta-globin promoter DNA methylation. *Blood* **110,** 1343–1352 (2007).
20. Tan, S. *et al.* Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci USA* **100,** 11997–12002 (2003).
21. Snowden, A. W., Gregory, P. D., Case, C. C. & Pabo, C. O. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Curr Biol* **12,** 2159–2166 (2002).
22. Xu, G. L. & Bestor, T. H. Cytosine methylation targetted to pre-determined sequences. *Nat Genet* **17,** 376–378 (1997).
23. McNamara, A. R., Hurd, P. J., Smith, A. E. F. & Ford, K. G. Characterisation of site-biased DNA methyltransferases: specificity, affinity and subsite relationships. *Nucleic Acids Res* **30,** 3818–3830 (2002).
24. Carvin, C. D., Parr, R. D. & Kladde, M. P. Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins. *Nucleic Acids Res* **31,** 6493–6501 (2003).
25. Smith, A. E. & Ford, K. G. Specific targeting of cytosine methylation to DNA sequences in vivo. *Nucleic Acids Res* **35,** 740–754 (2007).
26. Nomura, W. & Barbas, C. F. In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase. *J. Am. Chem. Soc.* **129,** 8676–8677 (2007).
27. Li, F. *et al.* Chimeric DNA methyltransferases target DNA methylation to specific DNA

sequences and repress expression of target genes. *Nucleic Acids Res* **35,** 100–112 (2007).
28. Rivenbark, A. G. *et al.* Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7,** 350–360 (2012).
29. Siddique, A. N. *et al.* Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. *Journal of Molecular Biology* **425,** 479–491 (2013).

Chapter 5

Discussion and Conclusions

Morgan L. Maeder

Rapid advances in the technologies for targeted genome editing have pushed this field to the forefront of molecular biology and biomedical research. As recently as a decade ago, the ability to engineer highly specific DNA-binding domains was the limiting factor in the field, and was a capability limited to only very few select zinc finger engineering labs. Recent developments, however, such as the advent of the TALE and CRISPR/Cas systems, have practically eliminated this hurdle and have opened up the field to pursue the use and applications of this powerful technology. In this dissertation, I describe three important applications of targeted DNA-binding protein technology.

Correction of a disease-causing mutation in human iPS cells

In Chapter 2, I described one of the first demonstrations that genome editing technology can be used to directly correct a disease-causing mutation in iPS cells derived from patients. Using ZFNs engineered by OPEN, we corrected the sickle cell anemia mutation in two human iPS cell lines. Along with three other concurrently published papers, this demonstrates an important proof-of-principle for the eventual goal of applying this technology to gene therapy[1-3]. Importantly, of these four studies, only ours used ZFNs engineered on a publically available platform, while the ZFNs used in the other three papers were made by the proprietary method of Sangmo/Sigma. Sangamo has routinely demonstrated their ability to engineer high quality ZFNs, however, access to these proteins is severely limited by the high price tag and reach-through rights imposed by Sigma. We therefore felt that it was critical to demonstrate that ZFNs engineered by the OPEN method were able to function efficiently in the context of correcting a disease mutation.

While this study takes the first step in applying genome editing technology to gene therapy, there is still a long road ahead. While we demonstrate genetic correction of the sickle cell mutation, we were unable to determine functional correction because β-globin is not expressed in iPS cells. There are protocols for the differentiation of iPS cells along the erythroid lineage, however, production of mature blood cells is extremely inefficient. Additionally, those cells that are obtained fail to express adult β-globin, but instead express fetal γ-globin, which raises the concern that these cells do not faithfully mimic endogenous blood cells. A better approach could be to use the described ZFN correction strategy to modify the genome of hematopoietic stem cells (HSCs). In theory, HSCs could be corrected and then reintroduced into the patient, where they would populate the hematopoietic system and differentiate naturally into erythrocytes.

However, this approach of manipulating multipotent stem cells in culture and then reintroducing them into a patient raises several critical safety concerns due to the fact that the modified cell population has the capacity to differentiate into many cell types. Besides the question of how tissue culture conditions would affect this cell population, the issue of ZFN specificity becomes even more critical when one considers that any off-target effects could be propagated throughout the hematopoietic system. Consider the example of the Sangamo *CCR5* ZFNs that are currently in clinical trials for the treatment of HIV/AIDs. $CD4^+$ T-cells are isolated from HIV patients and treated with ZFNs targeting *CCR5* in order to knockout the receptor and mimic the naturally occurring delta32 mutation, which renders individuals resistant to infection by the most common strain of HIV. In this approach, any off-target mutagenesis by the ZFNs is limited to the terminally differentiated T-cell population and is unlikely to result in adverse

oncogenic transformation effects. However, there is currently a pre-clinical stage program to investigate the use of *CCR5* ZFNs in hematopoietic stem cells and research is being conducted into direct *in vivo* application of these ZFNs. These approaches raise ZFN specificity as an issue of the utmost importance because any off-target mutations would be transmitted throughout many different cell lineages.

Several groups have begun to investigate the genome-wide specificity of ZFNs, however, much work remains to be done. Two studies that aimed to profile the CCR5 ZFNs discovered largely non-overlapping sets of off-target sites, implying that these are merely subsets of a much larger spectrum of off-targets. Additionally, the specificity of other nuclease platforms, such as TALENs and RGNs, remain largely unexplored. Genome editing nucleases are a powerful tool and hold great promise for the field of gene therapy but it is critical to ascertain the genome-wide effects of these proteins and failure to sufficiently profile and understand the specificity of these enzymes could have significant consequences.

Engineered TALE-Activators

In Chapter 3, I described the delineation of a robust TALE-activator platform for targeted manipulation of gene expression. Engineered ZF-activators had already proven to be a robust and useful tool for activating target genes and the expansion of this technology to engineered TALEs should broaden the targeting range of engineered transcription factors. Additionally, this study provided an opportunity to explore the robustness of TALE monomer fusion proteins. The published literature presented a worryingly low success rate for engineered monomeric TALEs, which contrasted sharply with the high efficiency and activity of dimeric TALENs. This

large-scale test of TALE-activators enabled us to explore whether this was due to variables in the engineering and testing of these proteins or to an inherent flaw in the affinity or specificity of monomeric TALE proteins.

We demonstrated in this work that, with use of an appropriate architecture, engineered TALE-activators constitute a robust framework for enabling targeted gene activation. Using this technology we also described the first demonstration of targeted activation of a microRNA cluster. One of the most interesting findings of this work is that multiple TALE-activators can function synergistically to activate transcription and this result was corroborated by another concurrently published study[4]. It has been shown that endogenous transcription factors can function synergistically to regulate gene expression by simultaneously binding parts of the transcription machinery[5-7]. The ability to recapitulate this important aspect of transcriptional regulation using engineered transcription factors enables study of the complex mechanisms that govern gene regulation. Additionally, this method provides a potential approach for engineering synthetic biology systems to control endogenous genes, similar to recently developed genetic circuits that control engineered transgenes[8].

In addition to developing a robust platform for activation of endogenous genes, this work demonstrates that TALE monomers can function to efficiently confer sequence specificity upon effector domains. This finding encouraged us to pursue engineering of various other TALE fusion proteins, including fusions to transcriptional repressors and enzymes that catalyze addition or removal of histone modifications and DNA methylation.

Targeted DNA demethylation by engineered TALE-TET1 proteins

In Chapter 4, I describe the engineering of targetable DNA hydroxymethylases capable of effecting specific CpG demethylation in human cells. Zinc fingers had previously been fused to various DNA methyltransferases and these ZF-DNMTs were shown to induce targeted DNA methylation[9,10], however, site-specific CpG demethylation had yet to be explored. I describe in this work that both TALEs and ZFs can be joined to the catalytic hydroxymethylase domain of the TET1 protein and that these fusion proteins can direct targeted DNA demethylation. In addition, I show that these demethylation events can result in increased gene expression at the endogenous human *RHOXF2* and *HBB* loci. Both targeted methylation and demethylation provide powerful tools for studying the effect of DNA methylation and establishing causality between specific methylation events and gene expression. The ability to specifically demethylate target sequences also provides a potential therapeutic approach to treating the myriad of cancers that are caused by aberrant methylation and silencing of tumor suppressors.

While our engineered TALE-TET1 fusion proteins efficiently remove methyl groups from target CpGs, the activity is currently limited to CpGs located approximately 10-150 bps away from the TALE binding site. Although expression of some genes is dependent on the methylation status of a small number of CpGs, expression is more commonly affected by multiple CpGs across larger regions, such as CpG islands. For this reason, a critical area of future optimization of this platform will be to extend the range over which demethylation can be induced. Initial attempts to increase the range of activity by lengthening the linker separating the TALE array and the TET1 catalytic domain failed to extend the demethylation effect. One potential approach could be to fuse engineered TALE arrays to a chain of

dimerization domains, such as the iDimerize system from Clonetech, that will recruit TET1 fused to a dimerization partner (Figure 5.1A). In this way, it may be possible to recruit multiple TET1 proteins and thereby increase the local concentration of hydroxymethlase activity at the target locus. Another approach could be to use the CRISPR/Cas9 system. While the endogenous Cas9 protein functions as a nuclease, it has been shown that its catalytic function can be inactivated by introducing two amino acid changes[11]. It may be possible to fuse the TET1 catalytic domain to the catalytic-dead Cas9 and in this way use guide RNAs to recruit hydroxmethylase activity. Although the multiplexing capability of the CRISPR/Cas system has yet to be fully explored, the demonstration that concurrent use of two gRNAs was able to induce Cas9-mediated cleavage of both genomic sites[12] suggests that simultaneous use of multiple gRNAs might successfully recruit Cas9-Tet1 fusions across a larger genomic region (Figure 5.1B).

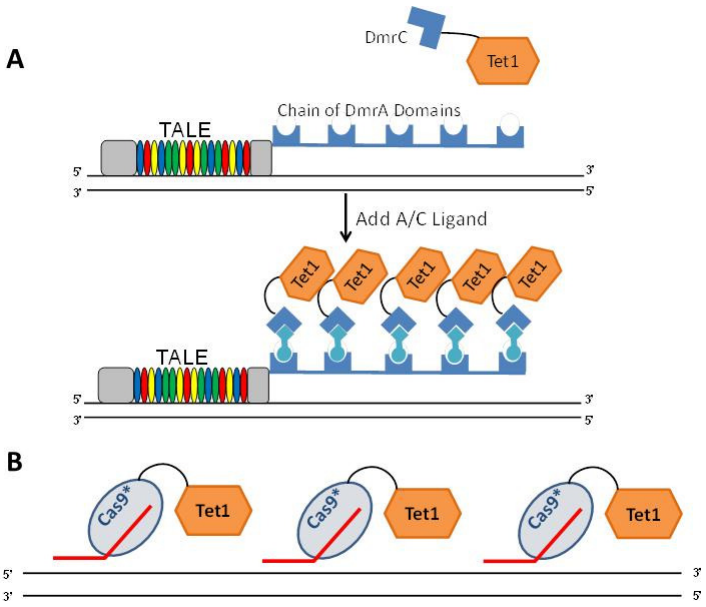


Figure 5.1. Strategies for Extending Demethylation. A) Schematic depicts engineered TALE array fused to a chain of DmrA dimerization domains. TET1 catalytic domain is fused to the heterodimeric partner DmrC. When small molecule dimerization ligand is added, multiple TET1 proteins are recruited to the locus. **B)** Schematic depicts Cas9-mediated recruitment of TET1 to multiple sites through the use of multiple gRNAs and catalytic-dead Cas9 (Cas9*) fused to TET1.

In addition to broadening the range of demethylation, future work should also focus on rendering the effect permanent. Our experiments show that CpGs are almost entirely re-methylated by two weeks post-transfection with TALE-TET1s and that gene expression

correspondingly decreases to its original level. This is not surprising as there are doubtless many control mechanisms at play in the cell to maintain correct methylation states once TALE-TET1 fusions are no longer expressed due to loss of the transiently transfected plasmids encoding these proteins. Not only DNA methylation, but also a complex network of histone modifications, is responsible for altering chromatin state and dictating gene expression. Therefore, it is possible that recruiting additional factors, such as histone methyltransferases that catalyze the methylation of H3K4, may serve to permanently alter chromatin state at the target locus and lock the gene in an active state. It will be interesting to explore concurrent targeting of DNA methylation and histone modifications and it is possible that this dual approach may serve to induce even more robust changes in gene expression.

Concluding Thoughts

In this dissertation I present my work on three applications of targeted DNA-binding domains. The development of my thesis work mirrors the progression in the field over the last five years. My previous work on developing the OPEN system to engineer zinc finger arrays enabled me to rapidly generate these proteins for targeted gene editing. As rapid advances in genome editing technology took place, I transitioned to the use of TALE DNA-binding domains. Elimination of the barrier to engineering DNA-binding proteins freed many researchers to pursue novel uses of these tools. After establishing the robustness of monomeric TALE proteins, I was interested in exploring the possibility of generating novel fusion proteins. I engineered fusions of TET1 to both TALEs and ZFs because I think it remains to be determined which of these platforms will be superior for genome-wide specificity and ease of delivery.

Likely, there are advantages and disadvantages to both technologies and these complementary techniques will both prove to be useful components of the epigenome engineering toolbox that will almost certainly be developed over the coming years. Additionally, new developments in the field, such as the CRISPR/Cas system, remain to be fully explored. The field continues moving forward at a rapidly increasing pace and the work described here represents my efforts to contribute to the development of new tools and to explore the exciting applications of these technologies.

References

1. Soldner, F. *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146,** 318–331 (2011).
2. Zou, J., Mali, P., Huang, X., Dowey, S. N. & Cheng, L. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* **118,** 4599–4608 (2011).
3. Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478,** 391–394 (2011).
4. Perez-Pinera, P. *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods* **10,** 239–242 (2013).
5. Carey, M., Lin, Y. S., Green, M. R. & Ptashne, M. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* **345,** 361–364 (1990).
6. Lin, Y. S., Carey, M., Ptashne, M. & Green, M. R. How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* **345,** 359–361 (1990).
7. Joung, J. K., Koepp, D. M. & Hochschild, A. Synergistic activation of transcription by bacteriophage lambda cI protein and E. coli cAMP receptor protein. *Science* **265,** 1863–1866 (1994).
8. Khalil, A. S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150,** 647–658 (2012).
9. Siddique, A. N. *et al.* Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. *Journal of Molecular Biology* **425,** 479–491 (2013).
10. Rivenbark, A. G. *et al.* Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7,** 350–360 (2012).
11. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
12. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).

Appendix 1

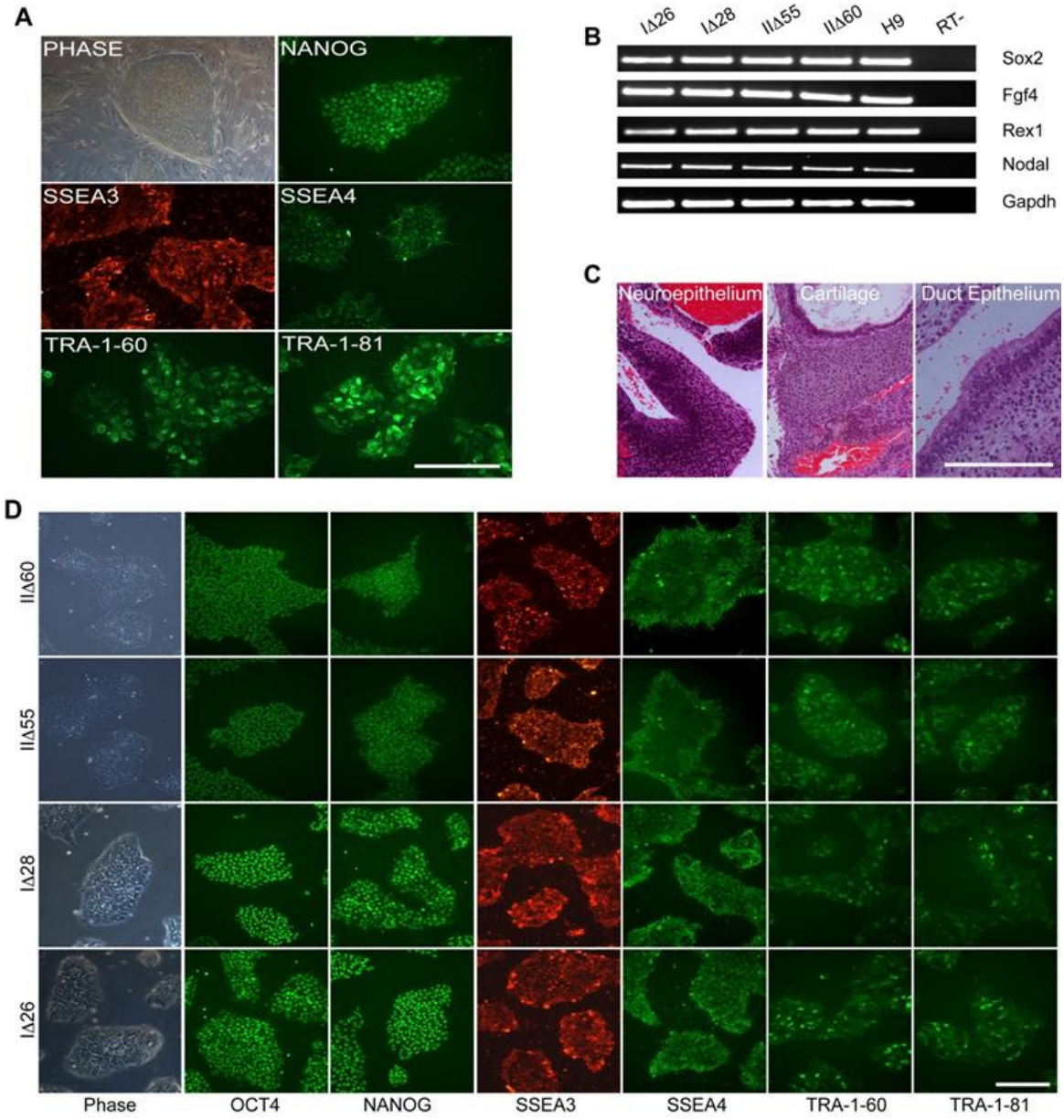
Supplementary Information for Chapter 2

In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases

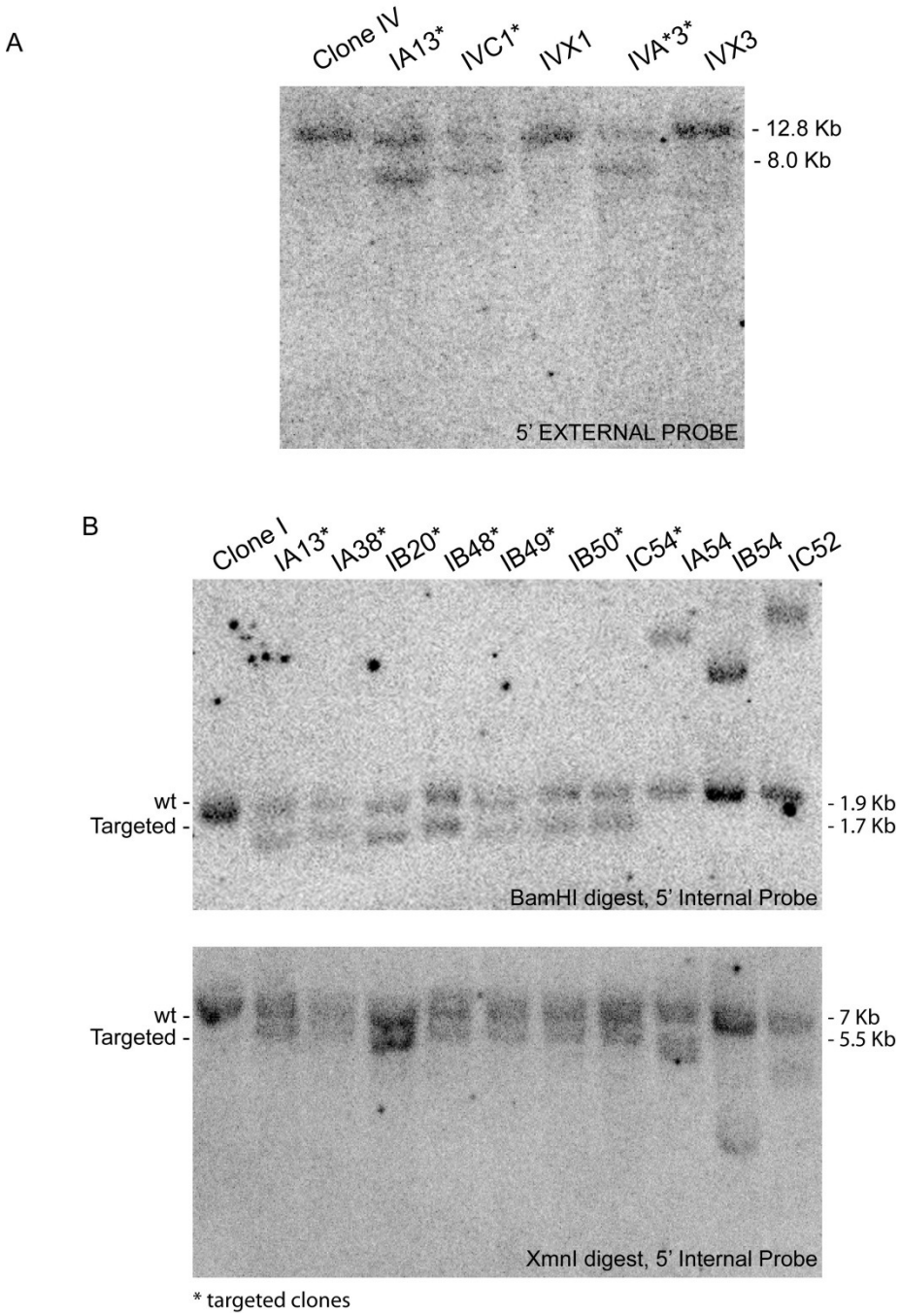
Vittorio Sebastiano,[1,2,*] Morgan L. Maeder,[3,4,*] James F. Angstman,[3] Bahareh Haddad,[1] Cyd Khayter,[3] Dana T. Yeo,[1] Mathew J. Goodwin,[3] John S. Hawkins,[1] Cherie L. Ramirez,[3,4] Luis F.Z. Batista,[5] Steven E. Artandi,[5] Marius Wernig,[1,#] & J. Keith Joung[3,4,6,#]

Morgan Maeder, with help from James Angstman, performed all Southern blots, PCR and sequencing to assess gene targeting and assay for off-targets. Vittorio Sebastiano performed all iPS cell clone verification including immunostaining for pluripotency markers and teratoma assays. Luis Batista performed TRAP assay.

Supplemental Figures and Tables



Supplementary Figure 2.1. Characterization of parental and integration-free parental iPS lines. A. Phase contrast image of iPS clone II grown on feeder cells and immunofluorescence images of feeder-free cultures. **B.** RT-PCR assays to assess expression of pluripotency-associated genes in looped-out iPS clones **C.** H&E staining of teratoma derived from iPS clone II showing tissues representative of all three germ layers. **D.** Immunofluorescence images of looped-out clones stained for pluripotency-associated transcription factors and surface markers in feeder-free cultures.



Supplementary Figure 2.2. Targeting of clone IV and analysis of donor template integration events. A. Southern blot using a 5' external probe on genomic DNA digested with *Pvu*II. A beta-globin allele that has not undergone gene targeting gives a 12.8kb band while a targeted allele gives an 8kb band due to the presence of a *Pvu*II site in the neomycin-resistance gene. Asterisks indicate gene-targeted clones while clones not marked with an asterisk are negative controls. **B.** Southern blots for assessing off-target integrations of the donor construct using a probe hybridizing to the 5' homology arm of the donor construct on genomic DNA digested with either *Bam*HI or *Xmn*I. The expected size for correct gene targeting events is 1.7kb for *Bam*HI digestion and 5.5kb for *Xmn*I digestion. Clones with asterisks were determined to be correctly targeted (as determined by PCR and Southern blot -- see Figure 2.3) while clones without asterisks were those that became negative after further expansion in culture.

A

ZFN Site 1

Beta globin	CCGTTACTGCCCTGTGGGGCAAGGT
Delta globin	CTGTCAATGCCCTGTGGGGCAAAGT
IA12	CTGTCAATGCCCTGTGGGGCAAAGT
IA13	CTGTCAATGCCCTGTGGGGCAAAGT
IA38	CTGTCAATGCCCTGTGGGGCAAAGT
a-gamma globin	CTATCACAAGCCTGTGGGGCAAGGT
IA12	CTATCACAAGCCTGTGGGGCAAGGT
IA13	CTATCACAAGCCTGTGGGGCAAGGT
IA38	CTATCACAAGCCTGTGGGGCAAGGT
g-gamma globin	CTATCACAAGCCTGTGGGGCAAGGT
IA12	CTATCACAAGCCTGTGGGGCAAGGT
IA13	CTATCACAAGCCTGTGGGGCAAGGT
IA38	CTATCACAAGCCTGTGGGGCAAGGT

ZFN Site 2

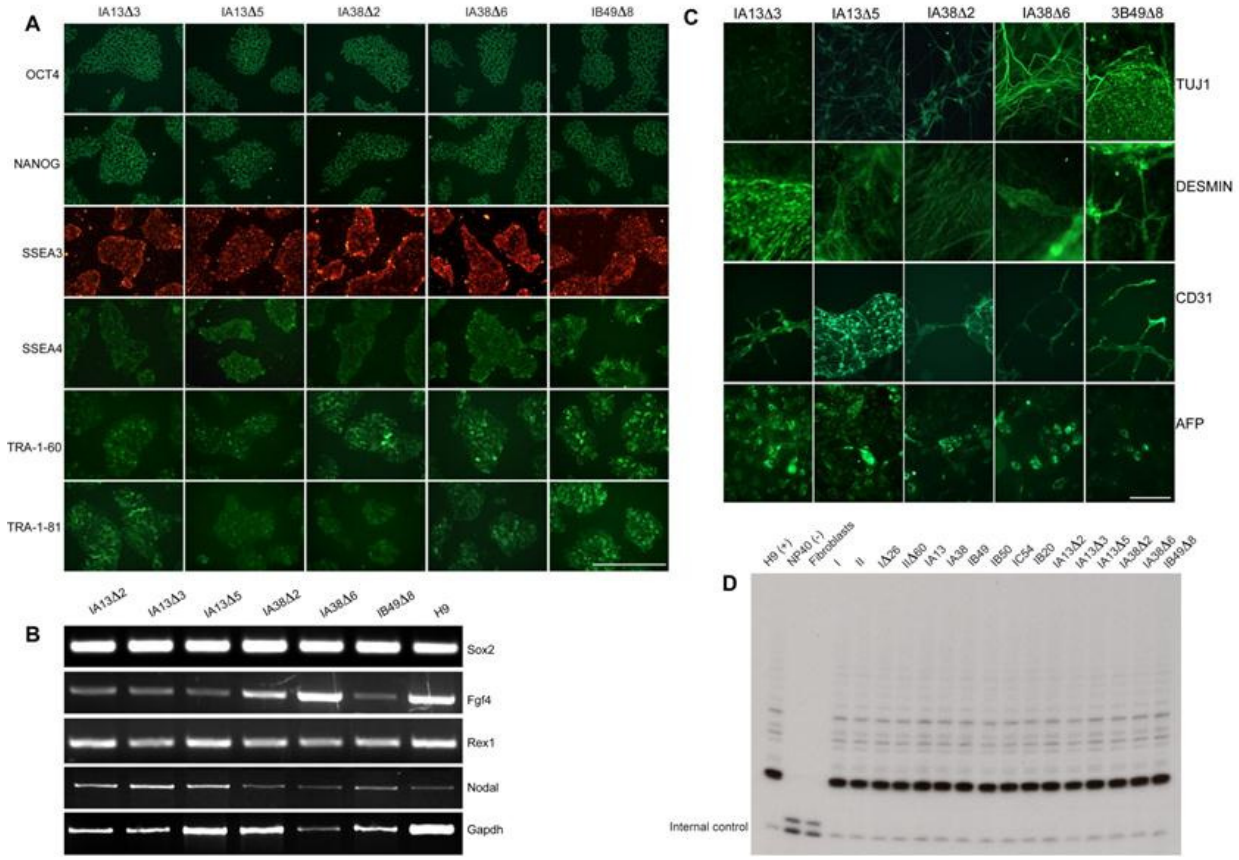
Beta globin sickle	ACCTGACTCCTGTGGAGAAGTCTGCC
Beta globin wt	ACCTGACTCCTGAGGAGAAGTCTGCC
Delta globin	ATCTGACTCCTGAGGAGAAGACTGCT
IB20	ATCTGACTCCTGAGGAGAAGACTGCT
IB48	ATCTGACTCCTGAGGAGAAGACTGCT
IB49	ATCTGACTCCTGAGGAGAAGACTGCT
IB50	ATCTGACTCCTGAGGAGAAGACTGCT
IC54	ATCTGACTCCTGAGGAGAAGACTGCT
IIB1	ATCTGACTCCTGAGGAGAAGACTGCT
a-gamma globin	ATTTCACAGAGGAGGACAAGGCTACT
IB20	ATTTCACAGAGGAGGACAAGGCTACT
IB48	ATTTCACAGAGGAGGACAAGGCTACT
IB49	ATTTCACAGAGGAGGACAAGGCTACT
IB50	ATTTCACAGAGGAGGACAAGGCTACT
IC54	ATTTCACAGAGGAGGACAAGGCTACT
IIB1	ATTTCACAGAGGAGGACAAGGCTACT
g-gamma globin	ATTTCACAGAGGAGGACAAGGCTACT
IB20	ATTTCACAGAGGAGGACAAGGCTACT
IB48	ATTTCACAGAGGAGGACAAGGCTACT
IB49	ATTTCACAGAGGAGGACAAGGCTACT
IB50	ATTTCACAGAGGAGGACAAGGCTACT
IC54	ATTTCACAGAGGAGGACAAGGCTACT
IIB1	ATTTCACAGAGGAGGACAAGGCTACT

B

ZFN A	
Beta globin	CGTTACTGC CCTGT GGGGCAAGG
chr12: 129185625-129185649	AGTTACTGC ACTTGA GGGGCAAGG
IA13	AGTTACTGC ACTTGA GGGGCAAGG
IA38	AGTTACTGC ACTTGA GGGGCAAGG
chr19: 17501464-17501488	CGTTACTGC CGGGAA GGGGGAAGG
IA13	CGTTACTGC CGGGAA GGGGGAAGG
IA38	CGTTACTGC CGGGAA GGGGGAAGG
chr5: 179126029-179126052	CTTTACTGC GGCGC GGGGCAAGG
IA13	CTTTACTGC GGCGC GGGGCAAGG
IA38	CTTTACTGC GGCGC GGGGCAAGG

ZFNs B+C	
Beta globin	CCTGACTCC TGTGGA GAAGTCTGC
chr1: 7194111-7194135	CCTGACTCC CCCCAG GTAGTCTGC
IB49	CCTGACTCC CCCCAG GTAGTCTGC
chr12: 111246882-111246905	CCTGACTCC TCCAG GAAGTCTCC
IB49	CCTGACTCC TCCAG GAAGTCTCC
chr18: 59582264-59582287	CCTAACTCC AGGAA GAAGTCTGC
IB49	CCTAACTCC AGGAA GAAGTCTGC
chr2: 197271577-197271601	CCTGACTCC CAATCT AAAGTCTGC
IB49	CCTGACTCC CAATCT AAAGTCTGC
chr20: 33451327-33451350	CCGGACTCC TTCCT GAAGTCTGC
IB49	CCGGACTCC TTCCT GAAGTCTGC
chr22: 39135349-39135372	CCTGACTCC AGGGC GTAGTCTGC
IB49	CCTGACTCC AGGGC GTAGTCTGC
chr4: 132834555-132834579	CCTGACTCC ATTCTT GAAGTCAGC
IB49	CCTGACTCC ATTCTT GAAGTCAGC
chr4: 188898306-188898330	CCTGACTCT GCCTTG GAAGTCTGC
IB49	CCTGACTCT GCCTTG GAAGTCTGC
chr5: 71315218-71315241	CCTGACTCC AGGGT GAACTCTGC
IB49	CCTGACTCC AGGGT GAACTCTGC

Supplementary Figure 2.3. Analysis of off-target effects. A. DNA sequencing of closely-related delta and gamma-globin paralogue genes in successfully targeted iPS cell clones. Segments of DNA sequence from the beta-globin locus are shown with zinc finger binding sites highlighted in yellow for both ZFN sites 1 and 2. Corresponding sequences from the delta-, a-gamma- and g-gamma-globin loci are also shown with bases differing from the beta-globin locus highlighted in red. DNA sequences of these paralogous loci in targeted clones are shown below. **B.** DNA sequencing of closely related, computationally identified potential off-target ZFN cleavage sites. The intended beta-globin target sites of ZFN pairs A, B, and C used for gene correction are shown with binding sites highlighted in yellow. Sites in the human genome that differ from the ZFN target site by only one bp are shown with the altered base in red with the genomic location for each site provided to the left. Unaltered sequences of these sites in successfullv corrected clones IA13. IA38 and IB49 are shown below each site.



Supplementary Figure 2.4. Characterization of integration-free gene corrected clones. A. Immunofluorescence images of iPS clones stained for pluripotency-associated transcription factors and surface markers in feeder-free cultures. **B.** RT-PCR analysis to assess expression of pluripotency-associated genes. **C.** Immunofluorescence images of iPS lines differentiated *in vitro*. **D.** Telomerase activity assays.

Supplementary Table 2.1. Characterization of parental iPS clones. Karyotype was determined by SKY-FISH; underlined: correct diploid chromosome spreads (chromosome number, sex-chromosomes, number of spreads shown in brackets); non-underlined: non clonal aberrations (number of observed chromosomes, extra or missing chromosomes are shown in bold and the number of spreads with the specified chromosome set are in brackets). *In vitro* differentiation: cells were differentiated via embryoid body (EB) formation for 8 days followed by culture of EB on gelatin-coated plates for 2 more weeks in media containing 20% FBS.

iPS LINE	KARYOTYPE	# OF INTEGRATIONS	IMMUNOSTAINING						IN VITRO DIFFERENTIATION				TERATOMA
			OCT 4	NANOG	SSEA3	SSEA4	TRA-1-60	TRA-1-81	TUJ1	CD31	DESMIN	AFP	
I	46, XX (13) 45, **-18** (1) 44, **-XX** (1) 43, **-8** (1)	3	+	+	+	+	+	+	+	+	+	+	+
II	46, XX (15) 47, **+3** (1) 45, **-1** (1) 44, **-18**, **-4** (1)	2	+	+	+	+	+	+	+	+	+	+	+

Supplementary Table 2.2. Characterization of gene corrected iPS clones.

Karyotype was determined by SKY-FISH; underlined: correct diploid chromosome spreads (chromosome number, sex-chromosomes, number of spreads shown in brackets); non-underlined: non-clonal aberrations (number of observed chromosomes, extra or missing chromosomes shown in bold, and the number of spreads with the specified chromosome set shown in brackets). *In vitro* differentiation: cells were differentiated via embryoid body (EB) formation for 8 days followed by culture of EB on gelatin-coated plates for 2 more weeks in media containing 20% FBS.

iPS LINE	KARYOTYPE	IMMUNOSTAINING						IN VITRO DIFFERENTIATION			
		OCT 4	NANOG	SSEA3	SSEA4	TRA-1-60	TRA-1-81	TUJ1	CD31	DESMIN	AFP
IΔ26	46, XX (13) 45, **-3** (1), **-5** (1)	+	+	+	+	+	+	n.d.	n.d.	n.d.	n.d.
IΔ28	46, XX (12) 47, **+14** (1), **+1** (1)	+	+	+	+	+	+	n.d.	n.d.	n.d.	n.d.
IIΔ55	46, XX (13)	+	+	+	+	+	+	n.d.	n.d.	n.d.	n.d.
IIΔ60	46, XX (20) 45, **-16** (1) 44, **-10**, **-11** (1)	+	+	+	+	+	+	n.d.	n.d.	n.d.	n.d.
IA13Δ2	46, XX (20) 45, **-14** (1) 44, **-X**, **-2** (1)	+	+	+	+	+	+	+	+	+	+
IA13Δ3	46, XX (18) 48, **+1**, **+3**, **+7** (1) 45, **-18** (1) 44, **-21**, **-4** (1)	+	+	+	+	+	+	+	+	+	+
IA13Δ5	46, XX (19) 47, **+X** (1), **+4** (2) 45, **-7** (1), **-X** (2)	+	+	+	+	+	+	+	+	+	+
IA38Δ2	46, XX (9) 45, **-13** (1) 44, **-1**, **-11** (1)	+	+	+	+	+	+	+	+	+	+
IA38Δ6	46, XX (18) 45, **-3** (1), **-7** (1), **-14** (1) 44, **-17**, **-20** (1)	+	+	+	+	+	+	+	+	+	+
IB49Δ8	46, XX (22) 45, **-1** (1), **-5** (1), **-18** (1) 44, **-12**, **-X** (1)	+	+	+	+	+	+	+	+	+	+

Supplementary Table 2.3. Characterization of integration-free parental and gene corrected iPS cell lines.
Cell line names, ZFN pairs used for targeting, and molecular characterization performed are shown for all gene targeted iPS cell clones described in this study. n.d. indicates that it was not possible to determine the allele targeted due to a mixed population of cells. "n.a." indicates that the assay or procedure was not performed.

	iPS Clones														
	Patient 1													Patient 2	
	IA12	IB20	IA5	IA13	IA38	IA54	IB48	IB49	IB50	IB54	IC43	IC52	IC54	IVA3	IVC1
ZFN Pair	A	B	A	A	A	A	B	B	B	B	C	C	C	A*	C
Allele targeted	sickle	sickle	non-sickle	sickle	sickle	n.d.	sickle	sickle	sickle	n.d.	n.d.	n.d.	sickle		
Positive by PCR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Positive by PCR after Expansion	n.a.	+	n.a.	+	+	-	+	+	+	-	-	-	+	n.a.	n.a.
Positive by Southern	n.a.	+	n.a.	+	+	-	+	+	+	-	n.a.	-	+	+	+

Supplemental Methods

RT-PCR

RNA was extracted using the RNA-Easy Mini Kit (Qiagen) according to the manufacturer's instructions. Reverse transcription was performed using Superscript III (Invitrogen) according to the manufacturer's instructions.

Identification and Analysis of Off-target sites

Human genome sequence (build 37.61) was scanned for all potential ZFN target sites that contain a single mismatch relative to the intended target site and with a spacer length of 5 or 6bp between the half-sites bound by each ZFN monomer. Genomic loci harboring these sites, as well as delta and gamma-globin genes were amplified by PCR from genomic DNA isolated from successfully targeted clones. The resulting PCR fragments were cloned into pCR4-TOPO vector using an Invitrogen TOPO-TA cloning kit and transformed into *E. coli*. For each locus, at least 6 clones were sequenced to assess whether ZFN-induced mutations were present in the two alleles.

Conditions for PCR Analysis

PCR was performed using either High Fidelity Platinum Taq (Invitrogen) or High Fidelity Accuprime (Invitrogen) according to the manufacturer's instructions. 100-200 ng of genomic DNA template was used in all reactions.

<u>Amplification for identification of NHEJ-mediated indel mutations:</u>

Forward β-globin primer: 5'-CCCTAGGGTTGGCCAATCTACTCC-3'

Reverse β-globin primer: 5'-CAGCCTAAGGGTGGGAAAATAGACC-3'

Amplification conditions: 95°C for 5 minutes; 25 cycles of 95°C for 30 seconds, 61°C for 30 seconds, and 68°C for 45 seconds; final extension at 68°C for 5 minutes. The expected amplicon size is 363bps.

PCR detection of gene targeting events (5' junction):

Forward β-globin primer: 5'- GATAGTCACACTTTGGGTTGTAAG-3'

Reverse Puro^R Primer: 5'- GTGGGCTTGTACTCGGTC-3'

Reverse PGK Primer: 5'- TACCGGTGGATGTGGAATGT-3'

Amplification conditions: 95°C for 5 minutes; 30 cycles of 95°C for 30 seconds, 52°C (Puro^R primer) or 58°C (PGK primer) for 30 seconds, 68°C for 2 minutes; final extension at 68°C for 7 minutes. The expected amplicon sizes are 1.6 kbs (Puro^R primer) or 1.2kbs (PGK primer).

PCR detection of gene targeting events (3' junction):

Forward Puro^R Primer: 5'- CCCGCAACCTCCCCTTCTAC-3'

Forward Neo^R Primer: 5'- AGAGGCTATTCGGCTATGACTG -3'

Reverse β-globin primer: 5'- GGTGCAAAGAGGCATGATACATTGTATC-3'

Amplification conditions: 95°C for 5 minutes; 30 cycles of 95°C for 30 seconds, 56°C for 30 seconds, 68°C for 2 minutes; final extension at 68°C for 7 minutes. The expected amplicon sizes are 1.3kbs (Puro^R primer) or 2kbs (Neo^R primer).

Amplification of gene targeted β-globin sequence:

Forward β-globin primer: 5'- TAGAAGACCTTTTCCCCTCCTACCC-3'

Reverse β-globin primer: 5'- CAAATGCAAAATTACCCTGATTTGGT-3'

Amplification conditions: 95°C for 5 minutes; 35 cycles of 95°C for 30 seconds, 62°C for 30 seconds, 68°C for 4 minutes; final extension at 68°C for 7 minutes. The expected amplicon sizes

are 1.9kbs for β-globin that has not undergone gene targeting and 3.5kbs for gene targeted β-globin.

PCR amplification for identification of potential delta-globin gene mutations:

Forward δ-globin primer: 5'- CTGAGTCAAGACACACATGACAGAAC-3'

Reverse δ-globin primer: 5'- GAGAAAAGTGAAGCATCTCCTGG-3'

Amplification conditions: 95°C for 5 minutes; 25 cycles of 95°C for 30 seconds, 61°C for 30 seconds, 68°C for 45 seconds; final extension at 68°C for 5 minutes. The expected amplicon size is 702 bps.

PCR amplification for identification of potential a-gamma- and g-gamma-globin gene mutations:

Forward γ-globin primer: 5'- CGGCTGACAAAAGAAGTCCTGGTATC-3'

Reverse γ-globin primer: 5'- CAAATCCTGAGAAGCGACCTGG-3'

Amplification conditions: 95°C for 5 minutes; 25 cycles of 95°C for 30 seconds, 63°C for 30 seconds, 68°C for 45 seconds; final extension at 68°C for 5 minutes. The expected amplicon size is 570 bps.

PCR amplification of other off-target loci:

off-target #1 forward primer: 5'- tggacagaggagacgagggtgag -3'

off-target #1 reverse primer: 5'- gttggtaagccctggccaacacc -3'

off-target #2 forward primer: 5'- ccgttcaggacgtctccagggtc -3'

off-target #2 reverse primer: 5'- agggacggggaacggaagagaag -3'

off-target #3 forward primer: 5'- acagccacttcctgcctcactcc -3'

off-target #3 reverse primer: 5'- aaagaaggctgtcgggccacctg -3'

off-target #4 forward primer: 5'- ggttactactcccatgtggccctc -3'

off-target #4 reverse primer: 5’- gattgcaccactgcactccagcc -3’

off-target #5 forward primer: 5’- agaccctggagccttgtccttcc -3’

off-target #5 reverse primer: 5’- tggttcctgccctgacctctatacc -3’

off-target #6 forward primer: 5’- agctggtattagaggcgcccacc -3’

off-target #6 reverse primer: 5’- ggcagtggggaggggacatttatc -3’

off-target #7 forward primer: 5’- ccctctcatctgcactgctacaagac -3’

off-target #7 reverse primer: 5’- agtgagccaggagagaccctgtc -3’

off-target #8 forward primer: 5’- acattaccccactccccactccc -3’

off-target #8 reverse primer: 5’- ccctgggtgccaggaaatttggc -3’

off-target #9 forward primer: 5’- tgtctcctcgctgaaggtggacg -3’

off-target #9 reverse primer: 5’- ggttccaggatggaagtgccaagc -3’

off-target #10 forward primer: 5’- tctgcgggtttttcatcccgtgg -3’

off-target #10 reverse primer: 5’- accaatcagcgctctgtaaaacacacc -3’

off-target #11 forward primer: 5’- cacaagtgtacctggcccacagc -3’

off-target #11 reverse primer: 5’- gccagatgcagacacattctgagtcac -3’

off-target #12 forward primer: 5’- actggggctcaggcttacaatgc -3’

off-target #12 reverse primer: 5’- ctgtgcccgtgagggaatgcttg -3’

Amplification conditions: 95°C for 5 minutes; 30 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 68°C for 45 seconds; final extension at 68°C for 5 minutes. The expected amplicon size is 500-700 bps.

<u>PCR to detect DNA encoding reprogramming factors:</u>

Forward WPRE primer: 5’- ACGCTGCTTTAATGCCTTTG-3’

Reverse WPRE primer: 5'- GAGATCCGACTCGTCTGAGG-3'

Amplification conditions: 95°C for 5 minutes; 30 cycles of 95°C for 30 seconds, 63°C for 30 seconds, 68°C for 45 seconds; final extension at 68°C for 5 minutes. The expected amplicon size is 470 bps for clones containing sequence encoding the four reprogramming factors.

PCR to verify undifferentiated state of iPS cells:

Forward Sox2: 5'-GGGAAATGGGAGGGGTGCAAAAGAGG-3'

Reverse Sox2: 5'-TTGCGTGAGTGTGGATGGGATTGGTG-3'

Forward Rex1: 5'-CAGATCCTAAACAGCTCGCAGAAT-3'

Reverse Rex1: 5'-GCGTACGCAAATTAAAGTCCAGA-3'

Forward Fgf4: 5'-CTACAACGCCTACGAGTCCTACA-3'

Reverse Fgf4: 5'-GTTGCACCAGAAAAGTCAGAGTTG-3'

Forward Nodal: 5'-GGGCAAGAGGCACCGTCGACATCA-3'

Reverse Nodal: 5'-GGGACTCGGTGGGGCTGGTAACGTTTC-3'

Amplification conditions: 95°C for 5 minutes, 30 cycles of 95°C for 30 seconds, 63°C for 30 seconds (Nodal and Sox2); 58°C (Fgf4); 57°C (Rex1); 60°C (Gapdh), 68°C for 45 seconds, followed by extension at 68°C for 5 minutes.

Southern Blots

All restriction enzymes used for probe preparation and digestion of genomic DNA were obtained from New England Biolabs.

c-Myc Probe: A 664 bp probe was excised from the STEMCCA vector using *Pvu*II and isolated from an agarose gel. For Southern blots using this probe, 15 µg of genomic DNA was digested overnight with either *BamH*I or *Nsi*I.

WPRE Probe: A 526 bp probe was excised from the STEMCCA vector using *Nae*I and *Cla*I and isolated from an agarose gel. For Southern blots using this probe, 15 µg of genomic DNA was digested overnight with HF-*Nco*I restriction enzyme.

5' β-globin External Probe: A probe that hybridizes upstream of the 5' donor arm in the endogenous β-globin locus was constructed by using PCR to amplify a DNA fragment from genomic DNA using forward primer 5'- CCTGGCTCACAAGTACCATTG-3' and reverse primer 5'-TCTGGGGCAAGTAAGAGGAG-3'. The resulting PCR product was TOPO-cloned into the pCR4TOPO vector (Invitrogen) and verified by DNA sequencing. A 598 bp probe was excised with *EcoR*I and isolated from an agarose gel. For Southern blots using this probe, 15 µg of genomic DNA was digested overnight with HF-*Pvu*II restriction enzyme.

3' β-globin External Probe: A probe that hybridizes downstream of the 3' donor arm in the endogenous β-globin locus was constructed by using PCR to amplify a DNA fragment from genomic DNA using forward primer 5'- TTTTACGGCGAGATGGTTTC
-3' and reverse primer 5'-GCCTTGATTTCAGAATCTTGC-3'. The resulting PCR product was TOPO-cloned into the pCR4TOPO vector (Invitrogen) and verified by DNA sequencing. A 586 bp probe was excised with EcoRI and isolated from an agarose gel. For Southern blots using this probe, 15 µg of genomic DNA was digested overnight with HF-PvuII restriction enzyme.

5' β-globin Internal Probe: A 668bp probe that hybridizes within the 5' donor arm was excised from the beta-globin donor template using *Asc*I and *Avr*II and isolated from an agarose gel. For

Southern blots using this probe, 15 µg of genomic DNA was digested overnight with HF-*Pvu*II restriction enzyme.

Puromycin Probe: A 612bp probe that hybridizes to the puromycin-resistance gene was excised from the beta-globin donor template using *Pvu*II and *EcoN*I and isolated from an agarose gel. For Southern blots using this probe, 15 µg of genomic DNA was digested overnight with either HF-*Nco*I or *Avr*II restriction enzymes.

Screening ZFNs for Cleavage Activity in Mammalian Cells

Plasmids encoding ZFN pairs were transfected into either 293T cells (ZFNs to Site 1) or primary human fibroblasts from a sickle cell anemia patient (ZFNs to Site 2) by Amaxa nucleofection and sorted for GFP+ cells by FACS. Three days following transfection, genomic DNA was isolated from cells and PCR was performed to amplify the region surrounding the ZFN cleavage site (Supporting Information). The resulting PCR product was TOPO-cloned into pCR4TOPO vector (Invitrogen) according to manufacturer's instructions. Plasmid DNA was prepared from 96 colonies and the cloned PCR fragment was sequenced to identify indel mutations introduced by non-homologous end-joining repair.

Detection of telomerase activity by TRAP

Human fibroblasts and fully undifferentiated H9 ES cells and iPS cells (grown in matrigel) were lysed in NP40 buffer (25 mM HEPES-KOH, 150 mM KCl, 1.5 mM MgCl2, 10% glycerol, 0.5% NP40, and 5 mM 2ME [pH 7.5] supplemented with protease inhibitors) for 15–30 min on ice.Extracts clarified by centrifugation at 16,000 x g for 10 min were quantified by Bradford

assay. Standard TRAP reaction was performed with 78 ng of protein, following a modified protocol from the manufacturer (TRAPeze kit, Chemicon).

Immunostaining

Cells were fixed in 4% PFA for 10 minutes at room temperature and either incubated in blocking solution (2% FBS in PBS) or permeabilized in 0.2% Triton X-100 followed by incubation in blocking solution for detection of cell surface markers and transcription factors, respectively. The primary antibodies used in the study are the following: anti-OCT4 (Santa Cruz); anti h-Nanog (Cosmo Bio); anti SSEA-3 and anti SSEA-4 (Developmental Studies Hybridoma Bank); anti TRA-1-60 and anti TRA-1-81 (Millipore); anti CD31 (R&D Systems); anti-desmin (Thermo Scientific); anti-AFP (R&D Systems). All primary antibodies were diluted 1:200 in blocking solution and incubated overnight at room temperature. Secondary antibodies (Alexa-488 and Alexa 594, Invitrogen) were diluted 1:1000 and incubated for two hours at room temperature.

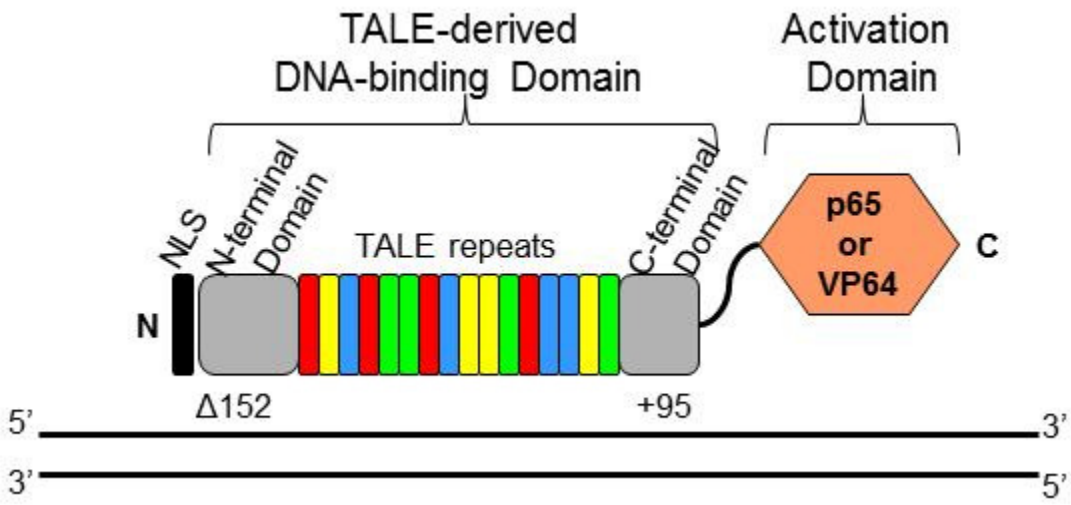
Appendix 2

Supplementary Information for Chapter 3

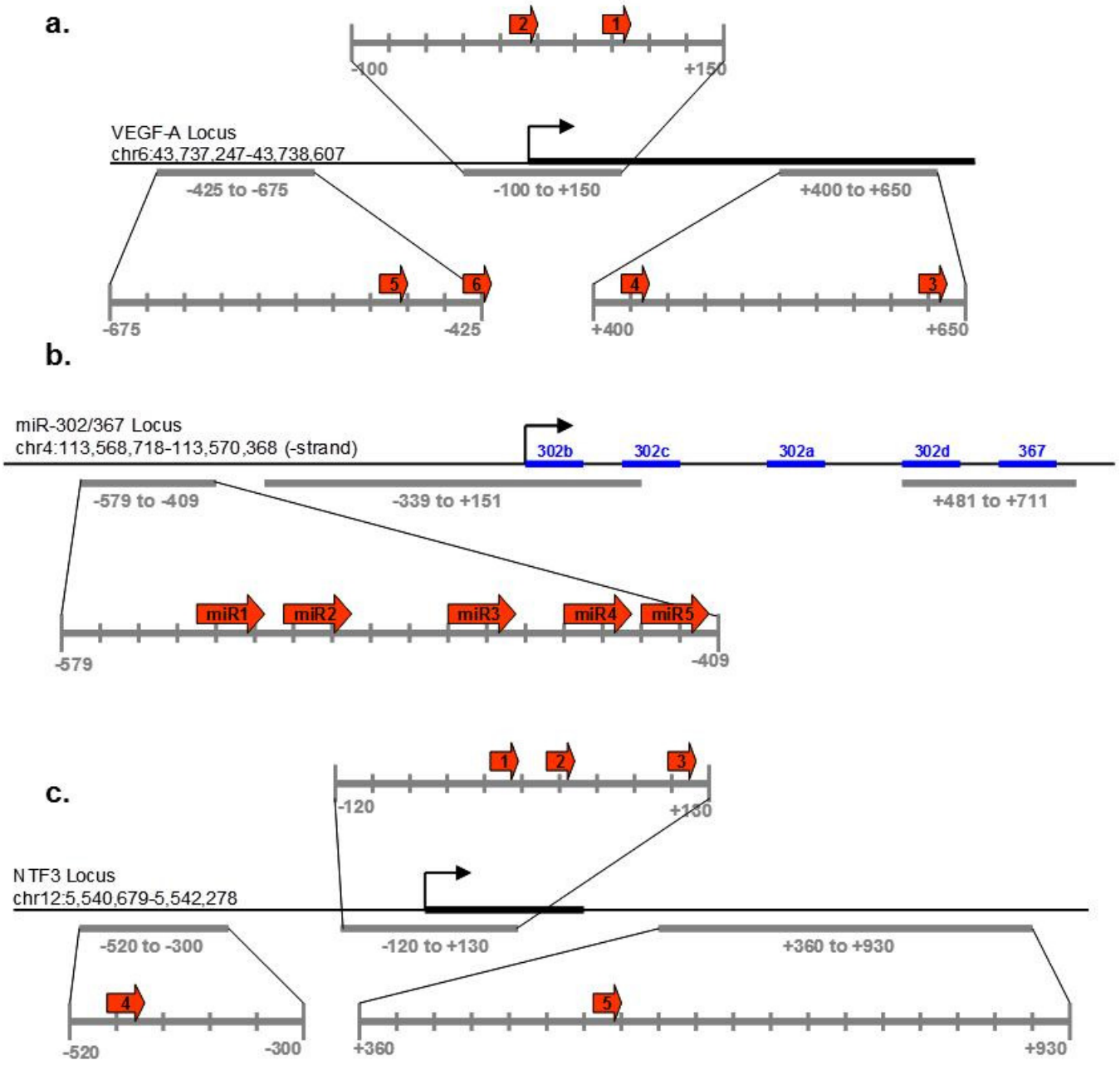
Robust and synergistic regulation of human gene expression using TALE-activators

Morgan L. Maeder[1,2], Samantha J. Linder[1,*], Deepak Reyon[1,3,*], James F. Angstman[1], Yanfang Fu[1,3], Jeffry D. Sander[1,3], J. Keith Joung[1,2,3]

Supplementary Figures and Tables



Supplementary Figure 3.1. Schematic of TALE-activator architecture used in this study. The TALE-activator architecture we used for our experiments is similar to one described by Rebar and colleagues.[1] These proteins contain the Δ152 N-terminal domain and the +95 C-terminal domain that flank the TAL effector repeat array as well as an N-terminal nuclear localization signal (NLS) and a C-terminal activation domain (either VP64 or p65).



Supplementary Figure 3.2. Schematic overview of TALE-activator binding sites within the (a) VEGF-A, (b) miR-302/367, and (c) NTF3 gene promoter regions. Thick black lines indicate exons, thin black lines indicate introns or promoter regions, and black arrows indicate the start site of transcription. Thick blue lines represent microRNAs. DNase I hypersensitive sites are indicated with grey bars and those we targeted are expanded with red arrows depicting precise locations of TALE-activator binding sites and orientations of the activators when bound on the DNA (the arrow indicates the direction of the protein from amino- to carboxy-terminus when bound to its target DNA site).

Supplementary Table 3.1. List of previously published TALE-activators and their reported activities on endogenous genes.[1-9]

Gene Targeted	Organism/Cell line	TALE length (# of repeats)	Activation Domain	Approximate Fold Activation	Reference	Architecture
NTF3	Human HEK293 cells	17.5	VP16	30	1	A
SOX2	Human 293FT cells	12.5	VP64	5.5	2	B
KLF4	Human 293FT cells	12.5	VP64	2.2		
OCT4	Human 293FT cells	12.5	VP64	no activation		
C-MYC	Human 293FT cells	12.5	VP64	no activation		
PUMA	Human HEK293T-Rex cells	17.5	VP16	1.5	3	C
IFNa1	Human HEK293T-Rex cells	19.5	VP16	3.1		
IFNb1	Human HEK293T-Rex cells	17.5	VP16	3.5		
FXN	Human 293FT cells	13.5	VP64	0.9 to 1.7	4	B
				1.1 to 1.6		
				1.0 to 1.6		
				1.1 to 2.0		
				1.1 to 1.4		
				1.7 to 3.1		
				1.1 to 1.5		
OSGIN2	Human U-2OS cells	18.5	VP64	4.8	5	D
ZC3H10	Human U-2OS cells	18.5	VP64	1.3		
ROCK1	Human HeLa cells	16.5	VP64	n.d.	6	E
CACNA1C	Human 293FT cells	16.5	VP64	5.5	7	B
				2.75		
				4.5		
				6		
				3		
				1.5		
				4		
				3.5		
OCT4	Mouse ES cells	17.5	VP16	4	8	F
	Mouse neural stem cells			30*		
Bs3	pepper plants	13.5	native AvrHah1 activation domain	n.d.	9	G

* Activation observed in the presence of VPA and/or 5-azadC

Architecture Key:
A = originally referenced in Miller et al., *Nature Biotech* 2011
B = originally referenced in Zhang et al., *Nature Biotech* 2011
C = originally referenced in Geissler et al., *PLoS ONE* 2011
D = originally described in Garg et al., *NAR* 2012
E = originally described in Huang et al., *Nature Biotech* 2011
F = originally described in Morbitzer et al., *NAR* 2011
G = originally described in Cermak et al., *NAR* 2011

Supplementary Table 3.2. List of target sites for TALE-activators tested in this study and their violations of computationally-derived guidelines previously described by Bogdanove and colleagues.[10] Guidelines are numbered as described in Supplementary Discussion with an X indicating violation. Asterisks indicate target sites for which we did not obtain functional TALE-activators.

TALE Name	Target site	Guidelines					Total Guideline Violations
		1	2	3	4	5	
VEGF1	TCGGGAGGCGCAGCGGTT					X	1
VEGF2	TTGGGGCAGCCGGGTAGC		X		X	X	3
VEGF3	TGGAGGGGGTCGGGGCTC				X	X	2
VEGF4	TGAGTGACCTGCTTTTGGG			X	X	X	3
VEGF5	TGAGTGAGTGTGTGCGTGT			X		X	2
VEGF6	TCACTCCAGGATTCCAATA			X	X		2
Ntf3-1	TTCTGTTCACGGGACTCA		X		X		2
Ntf3-2	TCCGAACAGCTCCGCGCA				X		1
Ntf3-3	TTCCCCTGCTGGGTAGTG		X		X	X	3
Ntf3-4	TACGCCTCAGACCTGATC				X		1
Ntf3-5	TCCCTCAATCTGGGAAAG				X		1
miR1	TGGAAGCAATCTATTTAT						0
miR2	TACATTTAACATGTAGAT						0
miR3	TAGAAACACAATGCCTTT						0
* miR4	TGGGAGCACTCATTGTTA				X	X	2
miR5	TAATCTATGCCATCAAAC			X	X		2
VEGF1-1	TTGGGGGTGACCGCCG		X		X	X	3
VEGF1-2	TTGGGGGTGACCGCCGGA		X		X	X	3
VEGF1-3	TTGGGGGTGACCGCCGGAGC		X		X	X	3
VEGF1-4	TTGGGGGTGACCGCCGGAGCGC		X		X	X	3
VEGF1-5	TTGGGGGTGACCGCCGGAGCGCGG		X		X	X	3
VEGF1-6	TTGGGGGTGACCGCCGGAGCGCGGCG		X		X	X	3
VEGF2-1	TCCCGCAGCTGACCAG				X	X	2
VEGF2-2	TCCCGCAGCTGACCAGTC				X		1
VEGF2-3	TCCCGCAGCTGACCAGTCGC				X	X	2
VEGF2-4	TCCCGCAGCTGACCAGTCGCGC				X	X	2
VEGF2-5	TCCCGCAGCTGACCAGTCGCGCTG				X	X	2
VEGF2-6	TCCCGCAGCTGACCAGTCGCGCTGAC				X	X	2
VEGF3-1	TACCACCTCCTCCCCG				X	X	2
VEGF3-2	TACCACCTCCTCCCCGGC				X	X	2
VEGF3-3	TACCACCTCCTCCCCGGCCG				X	X	2
VEGF3-4	TACCACCTCCTCCCCGGCCGGC				X		1
VEGF3-5	TACCACCTCCTCCCCGGCCGGCGG				X	X	2
VEGF3-6	TACCACCTCCTCCCCGGCCGGCGGCG				X	X	2

Supplementary Table 3.2 (continued)

TALE Name	Target site	Guidelines					Total Guideline Violations
		1	2	3	4	5	
VEGF4-1	TCCCCGGCCGGCGGCG				X	X	2
VEGF4-2	TCCCCGGCCGGCGGCGGA				X	X	2
VEGF4-3	TCCCCGGCCGGCGGCGGACA				X	X	2
VEGF4-4	TCCCCGGCCGGCGGCGGACAGT					X	1
VEGF4-5	TCCCCGGCCGGCGGCGGACAGTGG				X	X	2
VEGF4-6	TCCCCGGCCGGCGGCGGACAGTGGAC				X	X	2
VEGF5-1	TGGACGCGGCGGCGAG				X	X	2
VEGF5-2	TGGACGCGGCGGCGAGCC				X	X	2
VEGF5-3	TGGACGCGGCGGCGAGCCGC				X	X	2
VEGF5-4	TGGACGCGGCGGCGAGCCGCGG				X	X	2
VEGF5-5	TGGACGCGGCGGCGAGCCGCGGGC				X	X	2
VEGF5-6	TGGACGCGGCGGCGAGCCGCGGGCAG				X	X	2
VEGF6-1	TCCCAAGGGGGAGGGC				X	X	2
VEGF6-2	TCCCAAGGGGGAGGGCTC				X	X	2
VEGF6-3	TCCCAAGGGGGAGGGCTCAC				X	X	2
VEGF6-4	TCCCAAGGGGGAGGGCTCACGC				X	X	2
VEGF6-5	TCCCAAGGGGGAGGGCTCACGCCG				X	X	2
VEGF6-6	TCCCAAGGGGGAGGGCTCACGCCGCG				X	X	2
VEGF7-1	TCCGTCAGCGCGACTG				X	X	2
VEGF7-2	TCCGTCAGCGCGACTGGT					X	1
VEGF7-3	TCCGTCAGCGCGACTGGTCA				X	X	2
* VEGF7-4	TCCGTCAGCGCGACTGGTCAGC				X	X	2
VEGF7-5	TCCGTCAGCGCGACTGGTCAGCTG				X	X	2
VEGF7-6	TCCGTCAGCGCGACTGGTCAGCTGCG				X	X	2
VEGF8-1	TCCACTGTCCGCCGCC				X		1
VEGF8-2	TCCACTGTCCGCCGCCGG				X	X	2
VEGF8-3	TCCACTGTCCGCCGCCGGCC				X	X	2
VEGF8-4	TCCACTGTCCGCCGCCGGCCGG				X	X	2
VEGF8-5	TCCACTGTCCGCCGCCGGCCGGGG				X	X	2
VEGF8-6	TCCACTGTCCGCCGCCGGCCGGGGAG				X	X	2
VEGF9-1	TCCACCCCGCCTCCGG				X	X	2
VEGF9-2	TCCACCCCGCCTCCGGGC				X	X	2
VEGF9-3	TCCACCCCGCCTCCGGGCGC				X	X	2
VEGF9-4	TCCACCCCGCCTCCGGGCGCGG				X	X	2
VEGF9-5	TCCACCCCGCCTCCGGGCGCGGGC				X	X	2
VEGF9-6	TCCACCCCGCCTCCGGGCGCGGGCTC				X	X	2

Supplementary Discussion

Expanded the targeting range of monomeric TALE-activators

Cermak et al. originally proposed five guidelines for identifying optimal target sites of engineered dimeric TALENs.[9] These guidelines were computationally derived from data on the binding preferences of naturally occurring TAL effectors but were not prospectively tested experimentally. As summarized previously,[11] these five guidelines can be stated as follows:

1. The nucleotide just 5' to the first nucleotide in the TAL effector repeat array binding site should be a thymine.
2. The first nucleotide of the TAL effector repeat array binding site should not be a thymine.
3. The second nucleotide of the TAL effector repeat array binding site should not be an adenosine.
4. The 3' most nucleotide of the TAL effector repeat array binding site should be a thymine.
5. The base composition of the TAL effector repeat array binding site should not vary from the observed percent composition of naturally occurring binding sites by more than 2 standard deviations. The percent composition of naturally occurring TAL effector repeat array binding sites was determined to be: A = 31±16%, C = 37±13%, G = 9±8%, T = 22±10%. Therefore, the base composition of TALE binding sites should be: A = 0-63%, C = 11-63%, G = 0-25%, T = 2-42%.

In a large-scale study, our group recently demonstrated that highly active dimeric TALENs can be made for target sites that violate one or more of guidelines 2 through 5 (none of the sites we targeted violated guideline 1) and also showed that no significant correlation exists between the number of guideline violations and the activities of the engineered TALENs.[11]

More recently, Doyle et al. suggested that target site selection for monomeric TAL effector-based proteins should be limited by these same five guidelines.[10] The TALE-NT 2.0 web-based software tool (https://boglab.plp.iastate.edu/)[10] also applies these five guidelines in its default settings when choosing target sites for monomeric TAL effector repeat arrays used in TALE-activators.

The implementation of these guidelines has the effect of substantially limiting the targeting range of engineered monomeric TAL effector repeat arrays. For example, application of the five guidelines restricts the identification of a targetable 18 bp site (bound by a 16.5 TAL effector repeat array) to once in every 27 bps of random DNA sequence.

We used our data on 68 sites for which we were able to make active VP64 TALE-activators to test the importance of following five computationally-derived guidelines for target site choice. 65 of these 68 sites fail to meet one or more of these five guidelines with 56 of these sites violating two or more guidelines (note that all of the sites did meet the guideline requiring a 5' T) (Supplementary Table 3.2). Our ability to successfully obtain active TALE-activators for all 68 of these sites clearly demonstrates that there is no absolute requirement to follow at least four of the five design guidelines. We conclude that highly active monomeric TALE-activators can be made without meeting four of the five design guidelines. The ability to relax these restrictions improves the targeting range of TALE-activators by more than ten-fold –

for example, enabling proteins consisting of 16.5 TAL effector repeats to be made for a site once in every two bps of random DNA sequence, a more than 13-fold improvement in targeting range.

Web-based ZiFiT Targeter software for TALE-activator design

We have updated our publicly available Zinc Finger and TALE (ZiFiT) Targeter webserver to include tools designed to assist users interested in assembling monomeric TAL effector repeat arrays using our FLASH[11] assembly method. Support is also provided for making arrays using our lower-throughput REAL[12, 13] and REAL-Fast[13] methods because these approaches yield TAL effector repeat arrays that are identical in amino acid sequence to those made by FLASH. Our updated ZiFiT Targeter version 4.2 is currently available without registration at: *http://zifit.partners.org*. Identification of potential target sites using ZiFiT Targeter is performed by entering a sequence of interest into a query box. ZiFiT Targeter will, by default, identify TAL effector repeat arrays composed of 16.5 repeats that bind to sites 18 bp in length. Users can change this length constraint by entering a new value in the length input box.

Depending on the mode of assembly chosen (FLASH or REAL/REAL-Fast), ZiFiT will provide users with information about the names of plasmids required for assembly, and in the case of REAL/REAL-Fast assembly, a printable graphical guide. All plasmids required to practice REAL are available through the non-profit plasmid distribution service Addgene (http://www.addgene.org/talengineering/). The archive of 376 plasmids required to practice FLASH and REAL-Fast are available by request from the Joung Lab (http://www.TALengineering.org).

ZiFiT Targeter also provides tools to help users verify their TALE-activator plasmids after they have been assembled. As noted above, the DNA and amino acid sequences of TAL effector repeat arrays assembled by FLASH, REAL, and REAL-Fast are all identical. Users can download sequences for their specific engineered TAL effector repeat array using ZiFiT Targeter. Because alignment of TAL effector repeat array sequences can be challenging due to their highly repetitive nature, ZiFiT Targeter also provides a sequence alignment tool that attempts to align DNA sequence reads to the consensus sequence after anchoring sequences encoding the non-repetitive amino- and carboxy-terminal TAL effector-derived sequences that flank the repeat array.

Supplementary References

1. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148 (2011).

2. Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153 (2011).

3. Geissler, R. et al. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One* **6**, e19509 (2011).

4. Tremblay, J.P., Chapdelaine, P., Coulombe, Z. & Rousseau, J. TALE proteins induced the expression of the frataxin gene. *Hum Gene Ther* (2012).

5. Garg, A., Lohmueller, J.J., Silver, P.A. & Armel, T.Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* (2012).

6. Wang, Z. et al. An Integrated Chip for the High-Throughput Synthesis of Transcription Activator-like Effectors. *Angew Chem Int Ed Engl* **51**, 8505-8508 (2012).

7. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* **3**, 968 (2012).

8. Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* **40**, 5368-5377 (2012).

9. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).

10. Doyle, E.L. et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40**, W117-122 (2012).

11. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* (2012).

12. Sander, J.D. et al. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* **29**, 697-698 (2011).

13. Reyon, D., Khayter, C., Regan, M.R., Joung, J.K. & Sander, J.D. Engineering Designer Transcription Activator-Like Effector Nucleases (TALENs). *Curr Protoc Mol Biol.*, in press (2012).

Appendix 3

Supplementary Information for Chapter 4

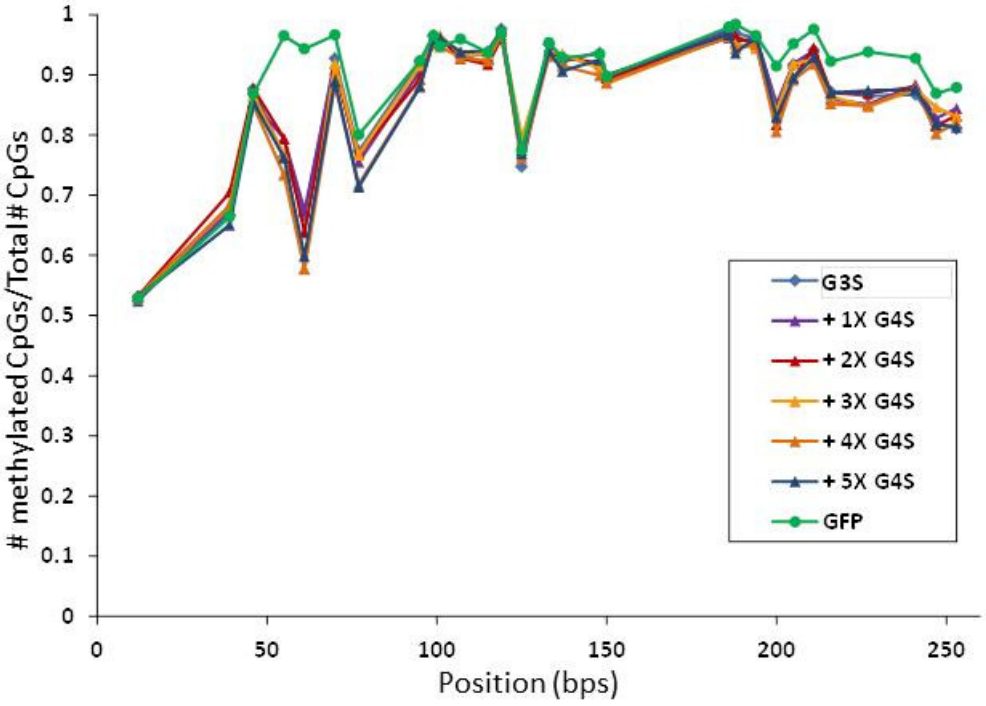
Targeted DNA Demethylation and Endogenous Gene Activation with Programmable TALE-TET1 Proteins

Morgan L. Maeder[1,2,3], James F. Angstman[1,2], Marcy E. Richardson[4], Samantha J. Linder[1,2], Vincent Cascio[1,2], Shengdar Q. Tsai[1,2,5], Quan Ho[1,2], Deepak Reyon[1,2,5], Jeffry D. Sander[1,2,5], Joseph F. Costello[6], Miles Wilkinson[4], J. Keith Joung[1,2,3,5,*]

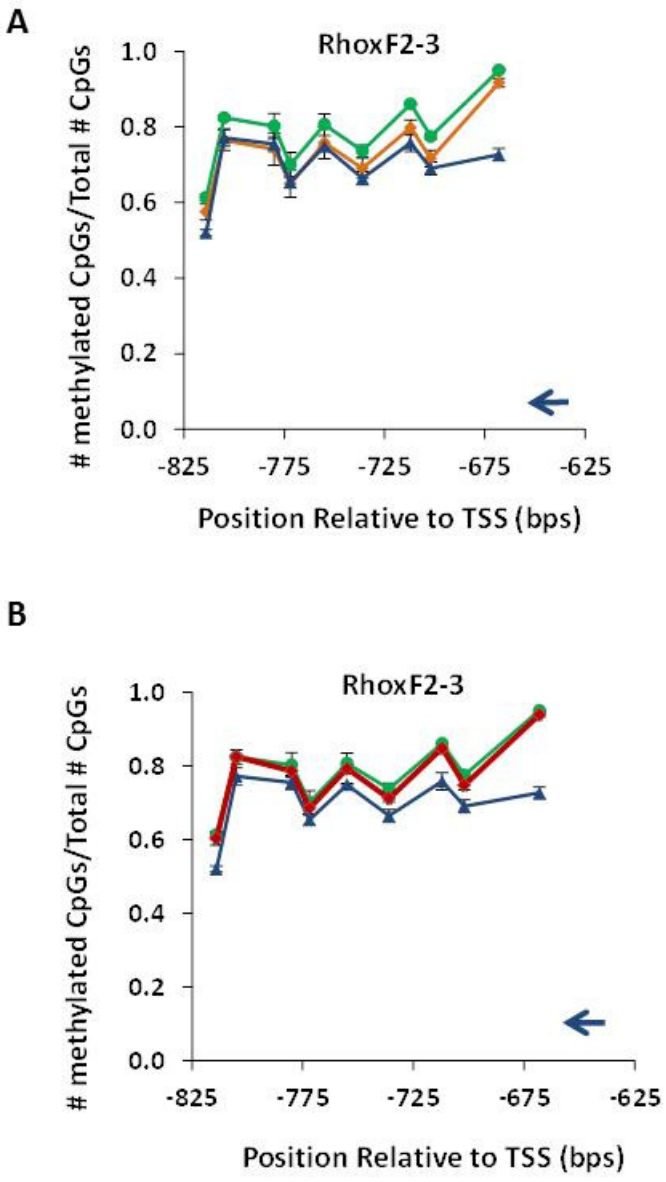
Supplementary Results

To test the feasibility of targeting TET1 activity with alternate DNA-binding domains, we engineered six-finger zinc finger (ZF) proteins targeted to two sites in the *KLF4* gene and to six sites in the *HBB* gene. ZFs were assembled using two-finger units according to the method originally described by Moore et al. (PNAS 2001) and similar to the ZF engineering approach used by Sangamo Biosciences (Doyon et al. 2008). The two-finger units used here were derived from three-finger proteins selected by the OPEN method and assembled six-finger proteins were fused to the catalytic domain of TET1. Of the two ZF-TET1 fusions targeted to KLF4, one failed to demethylate nearby CpGs while the other was able to induce demethylation of CpGs located 6 and 24bp downstream of the ZF binding site by 61% and 18%, respectively (Supplementary Figure 4.7). One ZF-TET1 fusion targeted to the -266 CpG in the HBB promoter failed to induce demethylation but all five fusions targeted to the region approximately 100 bp downstream of the transcription start site were able to demethylate target CpGs with efficiencies comparable to those observed with TALE-TET1 proteins (Supplementary Figure 4.8).

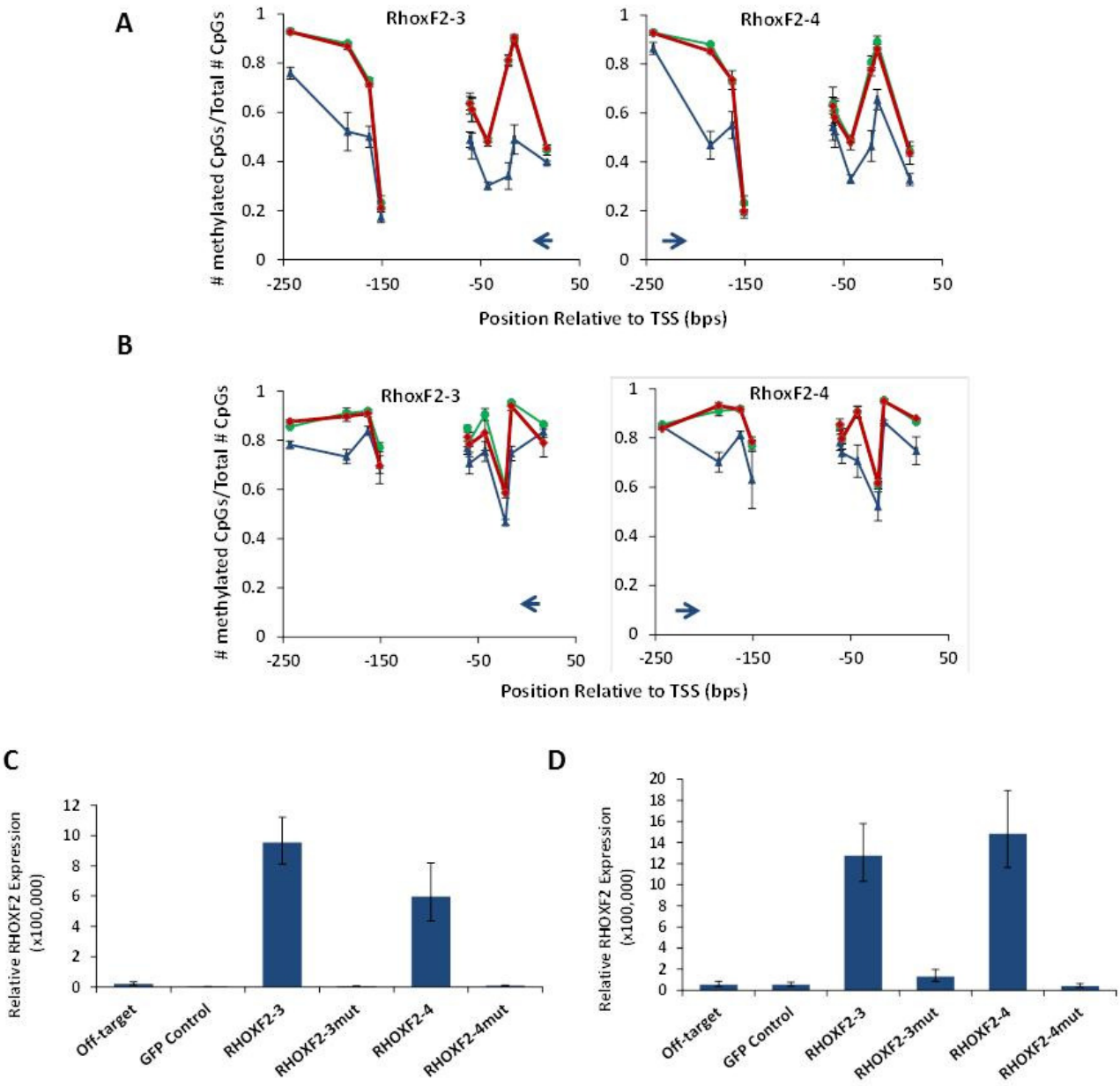
Supplementary Figures



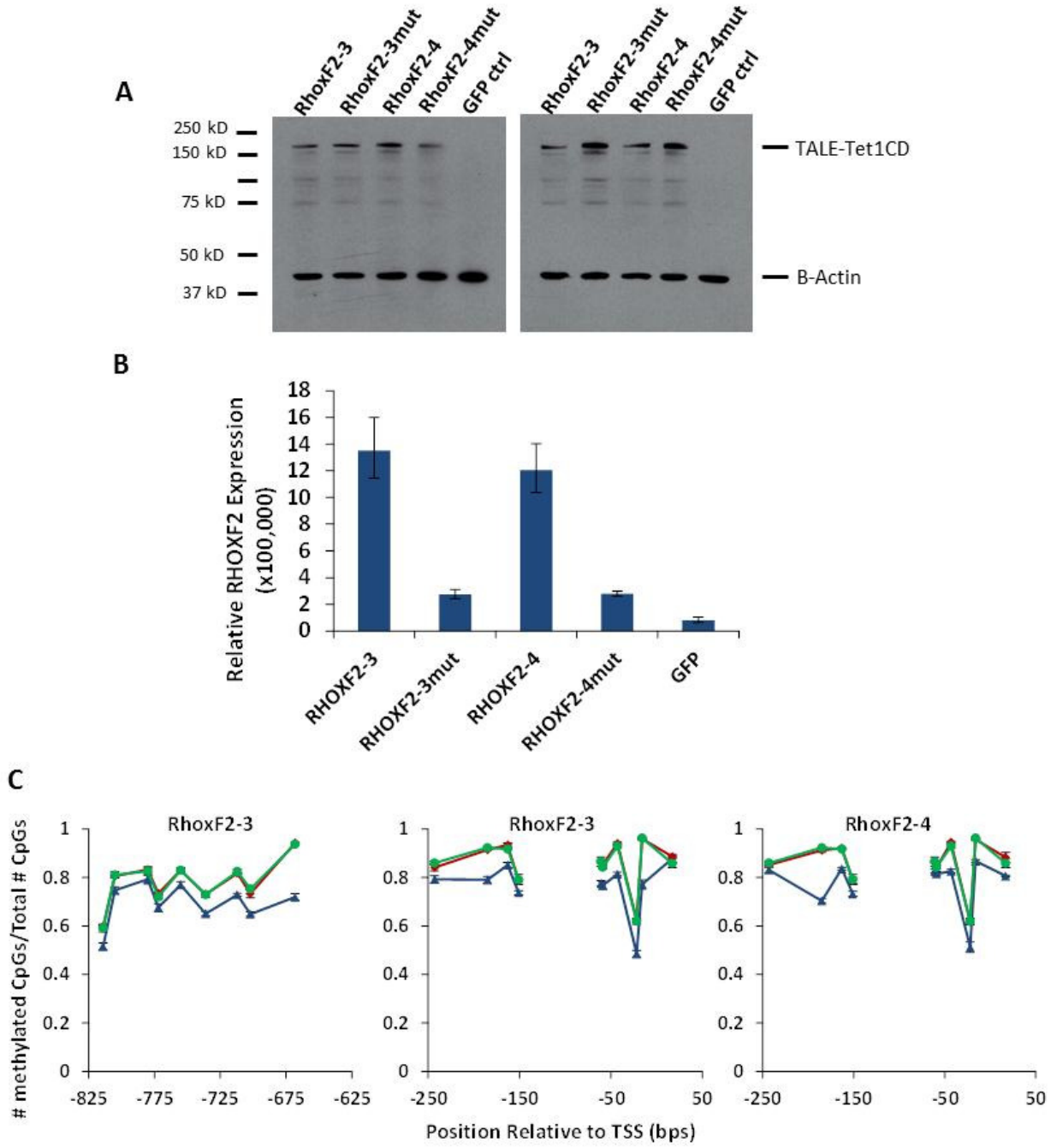
Supplementary Figure 4.1. TALE-Tet1 Extended Linkers. Graph shows the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in K562 cells transfected with KLF4-1-TALE-Tet1CDs with alternate linkers between the TALE and Tet1 domains or a GFP control (green).



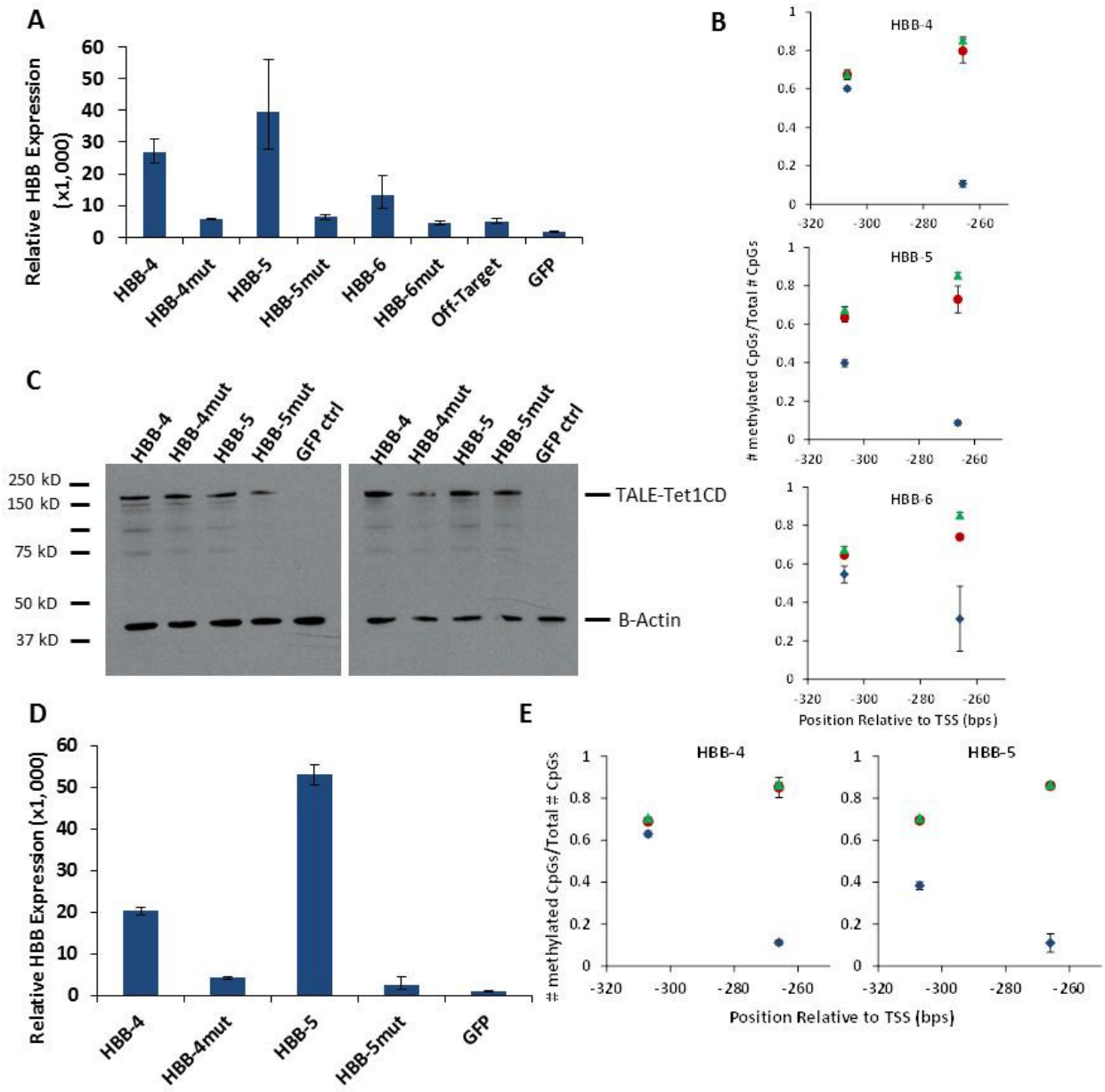
Supplementary Figure 4.2. RHOX-3 Activity at Alternate Binding Site. A) Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in 293 cells transfected with RHOX-3 TALE-Tet1CD (blue), an off-target TALE-Tet1CD control (orange) or a GFP control (green). Blue arrow depicts position of the TALE binding site. Error bars represent sem of three independent samples. **B)** Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in 293 cells transfected with RHOX-3 TALE-Tet1CD (blue), the catalytic mutant version (red) or a GFP control (green). Note that RHOX-3 and GFP samples are the same in both graphs and are depicted twice for ease of comparison.



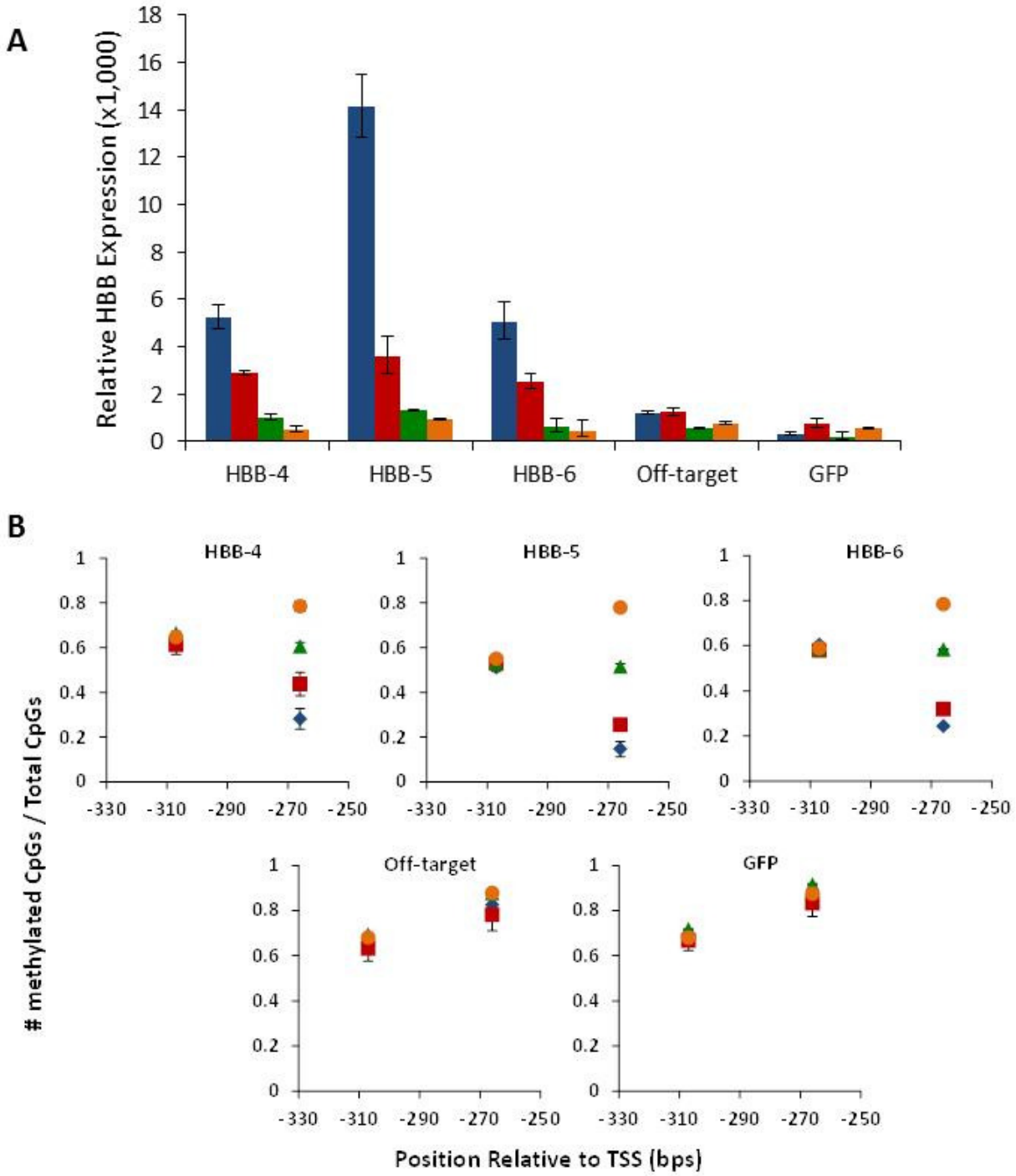
Supplementary Figure 4.3. TALE-Tet1CD Catalytic Mutants Targeted to the Human RHOXF2 Locus. A) Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in HeLa cells transfected with TALE-Tet1CD targeted to RHOXF2 (blue), the catalytic mutant version (red) or a GFP control (green). Note that GFP control is the same for both graphs. **B)** Same as A but in Hek293 cells. **C)** Expression levels of RHOXF2 mRNA, relative to β-Actin, in the same samples as A, as determined by qRT-PCR. Error bars represent sem of three independent samples, each assayed three times. **D)**Same as C but in Hek293 cells. Samples are the same as those used in B.



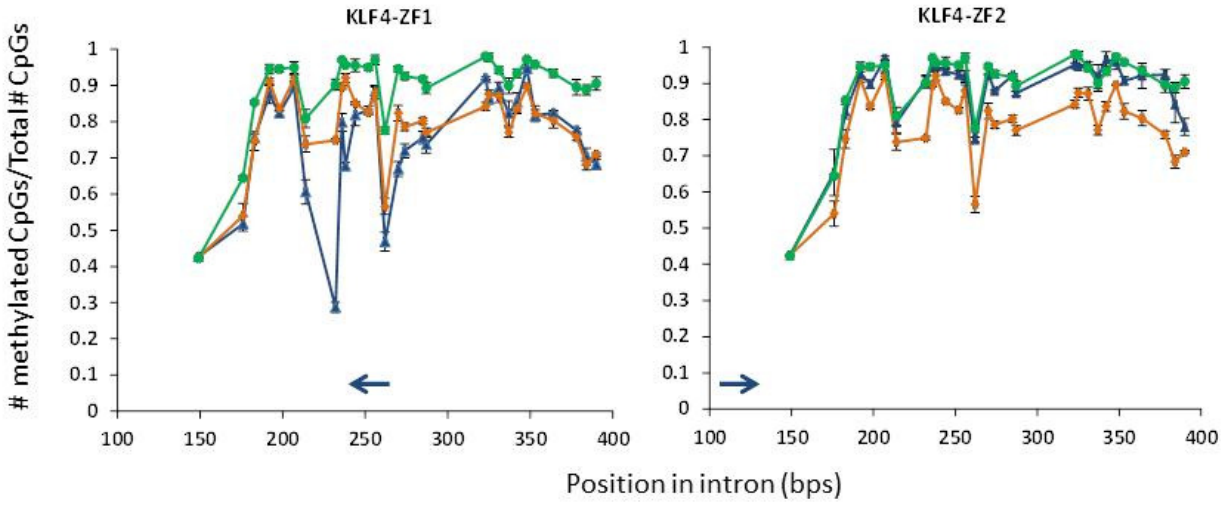
Supplementary Figure 4.4. Expression of TALE-Tet1CD Catalytic Mutants Targeted to the Human RHOXF2 Locus in 293s. A) Western blot analysis of TALE-Tet1CD and TALE-Tet1CDmut proteins expressed in Hek293 cells using anti-flag antibody. The two panels depict Western blots of independent sets of transfections. **B)** Expression levels of RHOXF2mRNA, relative to β-Actin, in the same samples used for Western blotting in A, as determined by qRT-PCR. Error bars represent sem of three independent samples, each assayed three times. **C)** Graph shows the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing in cells transfected with TALE-Tet1CD targeted to RHOXF2 (blue), the catalytic mutant version (red) or a GFP control (green). Note that GFP control is the same for all graphs. Same samples used in A and B .



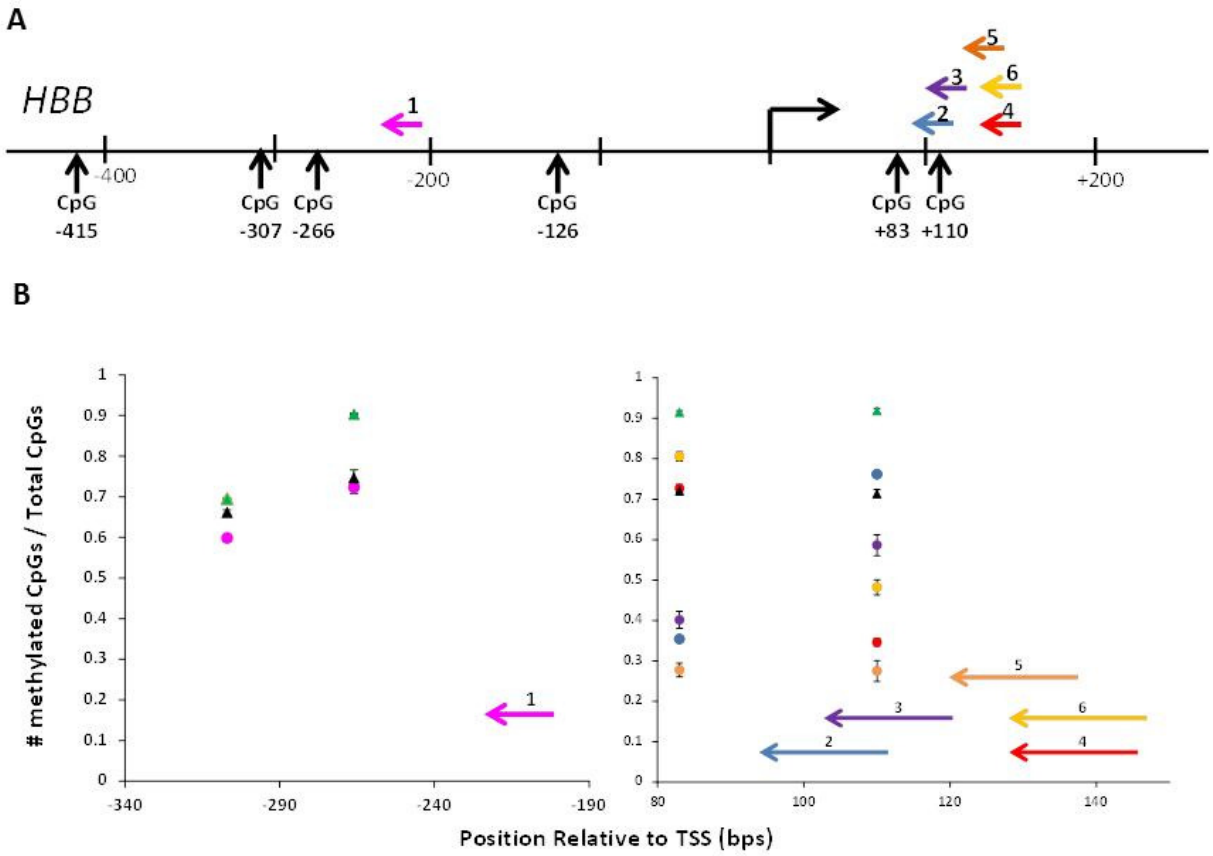
Supplementary Figure 4.5. TALE-Tet1CD Catalytic Mutants Targeted to the Human β-Globin Locus. A) Expression levels of HBB mRNA, relative to β-Actin, in K562 cells as determined by qRT-PCR. Error bars represent sem of three independent samples. **B)** Graph shows the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in cells transfected with TALE-Tet1CD (blue), TALE-Tet1CDmut (red) or a GFP control (green) (same samples as A). Note that GFP control is the same for all graphs. Error bars represent sem of three independent samples. **C)** Western blot analysis of TALE-Tet1CD and TALE-Tet1CDmut proteins expressed in K562 cells using anti-flag antibody. The two panels depict Western blots of independent sets of transfections. **D)** Expression levels of HBB mRNA, relative to β-Actin, in the same samples used for Western blotting in C, as determined by qRT-PCR. Error bars represent sem of three independent samples. **E)** Graph shows the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing in the same samples used in C and D .



Supplementary Figure 4.6. Demethylation of the Human β-globin Locus Over Time. A) Expression levels of HBB mRNA, relative to β-Actin, as determined by qRT-PCR assayed on day 4 (blue), day 7 (red), day 14 (green) and day 30 (orange) post-transfection. Error bars represent sem of three independent samples, each assayed three times. B) Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in K562 cells transfected with TALE-Tet1CDs targeted to the HBB locus, an off-target TALE-Tet1CD control or a GFP control assayed on day 4 (blue), day 7 (red), day 14 (green) and day 30 (orange) post-transfection. Error bars represent sem of three independent samples.



Supplementary Figure 4.7. ZF-Tet1CDs Targeted to KLF4. Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in K562 cells transfected with ZF-Tet1CDs (blue), an off-target ZF-Tet1CD (orange) or a GFP control (green). Blue arrows represent ZF binding sites and directionality. Note that off-target and GFP controls are the same in both graphs and are depicted twice for ease of comparison. Error bars represent sem of three independent samples.



Supplementary Figure 4.8. ZF-Tet1CDs Targeted to Human β-globin Locus. A) Schematic shows the HBB locus with CpGs indicated with black arrows and labeled relative to the transcription start site. ZFbinding sites are depicted as red arrows. **B)** Graphs show the number of methylated CpGs divided by the total number of CpGs, as determined by bisulfite sequencing, in K562 cells transfected with ZF-Tet1CDs targeted to the HBB locus (colored circles), an off-target ZF-Tet1CD control (black triangle) and a GFP control (green triangle). Position of ZF binding sites are depicted with colored arrows corresponding to the color of the circles on the graph. ZF targeted to the -266 CpG is depicted in the left panel and ZFs targeted to the +83 and +110 CpGs on the right. Error bars represent sem of three independent samples.

Supplementary Methods

Bisulfite PCR:

Standard Illumina multiplex adapters and indexes were added by PCR for KLF4 and HBB amplicons as described below.

KLF4 PCR 1 (For KLF4-1 TALE-TET1 and for ZF-TET1 samples)

Initial amplification and adding of Illumina multiplex adapters was performed on 1ul bisulfite-converted genomic DNA in a 40ul reaction using Kapa HiFi HotStart Uracil+ ReadyMix, a final concentration of 6nM for each initial primer and 30nM for each primer containing Illumina adapter and the following PCR conditions: 95°C for 5 minutes; 10 cycles of 98°C for 20 seconds, 68.5°C for 15 seconds, 72°C for 20 seconds; 32 cycles of 98°C for 20 seconds, 72°C for 25 seconds; 72°C for 5 minutes. PCR was purified with Agencourt AMpure XP PCR purification beads (Beckman Coulter). Approximately 2ng purified initial PCR was used as template for PCR with Illumina Multiplexing PCR primer 1.0 and one of 60 Illumina Index primers (see Supplementary Table 1) using 2X Phusion Hot Start Flex DNA polymerase (NEB) and the following PCR conditions: 98°C for 2 minutes; 10 cycles of 98°C for 20 seconds, 57°C for 30 seconds, 72°C for 30 seconds; 72°C for 5 minutes. PCR was purified with Agencourt AMpure XP PCR purification beads (Beckman Coulter), quantified on a QIAxcel (Qiagen) and equal molar amounts of each amplicon were combined.

KLF4 PCR 2 (For KLF4-2 TALE-TET1 samples)

Exactly as described for KLF4 PCR 1 with the exception that the initial PCR contained 1% DMSO and used the following conditions: 95°C for 5 minutes; 10 cycles of 98°C for 20 seconds, 66°C for 15 seconds, 72°C for 20 seconds; 32 cycles of 98°C for 20 seconds, 72°C for 25 seconds; 72°C for 5 minutes.

HBB PCRs 1 and 2

Exactly as described for KLF4 PCR 1 but with the following PCR conditions for initial PCR: 95°C for 5 minutes; 10 cycles of 98°C for 20 seconds, 63.2°C for 15 seconds, 72°C for 20 seconds; 32 cycles of 98°C for 20 seconds, 72°C for 25 seconds; 72°C for 5 minutes. Additionally, the final combined library was gel isolated from a polyacrylamide gel to remove primer dimers.

Standard Illumina multiplex adapters were ligated to RHOX bisulfite PCR amplicons and indexes were added by PCR as described below.

RHOXF2 PCR 1

Initial amplification was performed on 4ul bisulfite-treated genomic DNA in a 50ul reaction using Pyromark PCR kit (Qiagen) according to manufacturer's protocol with the following PCR conditions: 95°C for 15 minutes; 45 cycles of 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 30 seconds; 72°C for 10 minutes. PCR was purified with Agencourt AMpure XP PCR purification beads (Beckman Coulter). End repair was performed with a combination of

T4 DNA polymerase, Klenow DNA polymerase and T4 polynucleotide kinase (NEB) and A-tailing was performed with Klenow (3' to 5' exo-) and 200nM dATP (NEB). Standard Illumina multiplex adapters were ligated using T4 DNA ligase (NEB) and final product was purified using AMpure XP. Approximately 5-10ng purified adapter-ligated initial PCR was used as template for PCR with Illumina Multiplexing PCR primer 1.0 Illumina Multiplexing PCR primer 2.0 (see Supplementary Table 1) using Pyromark PCR kit (Qiagen) and the following PCR conditions: 95°C for 15 minutes; 6 to 10 cycles (depending on template concentration) of 94°C for 20 seconds, 65°C for 30 seconds, 72°C for 30 seconds; 72°C for 5 minutes. PCR was purified with Agencourt AMpure XP PCR purification beads (Beckman Coulter) and quantified on a QIAxcel (Qiagen). 1ng purified PCR was used as template for PCR with with Illumina Multiplexing PCR primer 1.0 and one of 60 Illumina Index primers (see Supplementary Table 1) using Pyromark PCR kit and the following PCR conditions: 95°C for 15 minutes; 10 of 94°C for 20 seconds, 57°C for 30 seconds, 72°C for 30 seconds; 72°C for 5 minutes. PCR was purified with Agencourt AMpure XP PCR purification beads (Beckman Coulter), quantified on a QIAxcel (Qiagen) and equal molar amounts of each amplicon were combined. The final combined library was gel isolated from a polyacrylamide gel to remove primer dimers.

RHOXF2 PCR 2

Exactly as described for RHOXF2 PCR 1 but with the following PCR conditions for initial amplification: 95°C for 15 minutes; 45 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 72°C for 30 seconds; 72°C for 10 minutes.